Enhancer Induced Changes In Chromatin Structure And Function

By

Jonathon D. Bennett

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

September 2014

Abstract

Gene regulation in mammals is a cornerstone issue in developmental biology. One example of gene regulatory elements are enhancers, which have been shown in transfection assays to be *cis*-regulatory sequences that activate transcription regardless of location and orientation to their gene promoters. However, no studies have examined *in vivo* whether the distance of a mammalian enhancer is critical for its function, and whether enhancers affect other genes beyond proximal regulatory domains. One significant mammalian enhancer is located within the immunoglobulin heavy chain (*IgH*) gene locus that spans a distance of 3 Mb on mouse chromosome 12. The *IgH* locus contains variable (V_H), diversity (D_H), and joining (J_H) gene segments, and constant region (C_H) exons. These *IgH* gene segments recombine early in bone marrow B-lymphocyte differentiation to yield functional antigen receptor genes. Coupled to this, it has been demonstrated that the Eμ intronic enhancer, located between the J_H-Cμ intron, has a significant role in gene regulation, including V(D)J recombination. The first project is to test whether the location of the Eμ enhancer is required for its function with respect to chromatin, transcription, and recombination. The second project is to determine whether the Eμ enhancer affects other genes in *cis* within the *IgH* locus, or elsewhere in the murine genome. For the first project, BAC transgenic mice were developed that have the Eμ enhancer in different locations and experiments assayed for changes in chromatin, transcription, and recombination as a result of the relocated Eμ enhancer. Our results demonstrate that regions of heterochromatin are changed to euchromatin based on the location of the Eμ enhancer, and that when the location of the Eμ enhancer is changed, unique transcripts and gene rearrangements are induced. For the

second project, genome-wide DNase Hypersensitivity sequencing was performed on samples with the Eμ enhancer present and absent. Several novel peaks associated with DNase-I sites within the *IgH* locus were identified, and some peaks were lost when the Eμ enhancer was absent. Therefore, the Eμ enhancer has other unique gene regulatory features in its proximal domain, and each of the unique DHS peaks may serve a gene regulatory purpose for the *IgH* locus that needs further exploration.

Mentor: Dr. Ranjan Sen

Reader: Dr. Joel Pomerantz

Acknowledgements

First, I thank the Immunology Training Program at Johns Hopkins School of Medicine for providing the opportunity to pursue advanced scientific training in the diverse field of immunology. I thoroughly enjoyed the coursework, retreats, seminars, and laboratory experience afforded by the faculty. It was through this program that I met my scientific mentor, Dr. Ranjan Sen, at the National Institutes of Health, National Institute of Aging. I am very thankful for the opportunity to work in his laboratory and receive his guidance and advice in scientific projects. I thoroughly enjoyed his work ethic and enthusiasm towards scientific discovery. He taught me how to overcome experimental problems and was patient throughout the process. I also want to thank all the members of the Sen Laboratory from 2009 – 2014. They all were critical in generating an enjoyable laboratory environment to pursue rigorous scientific studies.

Second, I want to thank my thesis committee members including Dr. Joel Pomerantz, Dr. Nan Ping Weng, Dr. Antony Rosen, Dr. Sebastian Fugmann, and Dr. Zhou Zhu. They gave valuable support and assisted with the progress of all the projects. In addition, they assisted with my growth in considering different scientific approaches to solve experimental problems.

Third, I want to thank my family, including those in Maryland, North Carolina, Pennsylvania, New York, and Oregon. The support of each member will be forever remembered, even as the family population both grew and diminished during these years of training. My father, Tom, tragically passed away while I was in Baltimore, and I wish he could have seen this body of research. He taught me there is "no off season." I learned this especially in the past 6 years of scientific training, and will continue this

theme into the next chapter of my life. Lastly, I thank my wife, Katie for her patience, support, encouragement, and care for our family. There were so many packed lunches, dinner drop-offs, washed clothes, notes of assistance, that a large part of my survival was dependent on her continual, faithful presence. She is a special person, and I am very fortunate and grateful to have married her in 2005.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 1: The Critical Importance of the Eμ enhancer

Introduction

One major beauty in all of life is the ability of jawed vertebrates and their immune systems to recognize and efficiently remove foreign antigens as they are presented on a vast repertoire of substances. Lymphocytes are integral to the adaptive immune system and are able to respond to individual antigens through the presence of antigen receptors present on their cell surface. Such specialized lymphocytes arise from a common lymphoid progenitor (CLP) in the bone marrow that gives rise to B, T, and Natural Killer (NK) cells. While NK cells lack antigen specificity, B and T cells both have monoclonal high affinity antigen receptors. Two key differences separate B and T cells: (1) B-cells differentiate within the bone marrow tissue and T-cells differentiate in the thymus, and (2) upon encounter with antigen, B-cells have the ability to differentiate into antibody producing plasma cells, follicular B cells, or memory B cells, whereas T-cells differentiate into several other sub-types, each with a variety of functions. Nonetheless, these two lymphocyte populations share one major characteristic.

Though different in tissue origination and effector function, immature B and T cells each have antigen-specific receptors that are formed by the process of V(D)J recombination. As puzzle pieces combine in a position-specific manner to produce a picture, so do the separate variable (V), diversity (D), and joining (J) gene segments rearrange and combine to produce a coding sequence that is able to generate diverse antigen receptors. In total, there are seven antigen-receptor loci in B and T cells. This includes the α, β, γ, and δ loci expressed in T-cells to produce the T-cell receptor, and the *IgH* and light (Igκ and Igλ) chain loci expressed in B-cells to produce a functional B-cell

receptor. Hundreds of gene segments are within the V(D)J loci, and it is the assembly and combination of these gene segments during B-cell development that generates the large range of sequences that encode unique antigen recognition motifs. While a comprehensive treatment of antigen loci rearrangement and their *cis*-regulatory elements includes the study of all seven antigen-receptor loci in B and T cells, the specific focus in this thesis is that of the *cis*-regulatory features that drive somatic recombination in pro-B cells on the mouse *IgH* chain locus, a 3 Mb region located on the long arm of chromosome 12 and approximately 5 Mb to the telomere.

In order for V(D)J recombination to take place, the site-specific recombination activating (RAG) nuclease, which corresponds to the lymphoid moiety of V(D)J recombinase, is needed to catalyze the precise cutting and pasting of individual V, D, and J gene segments, along with the ubiquitously expressed DNA repair factors from the non-homologous end joining pathway (Johnson et al., 2010). V(D)J recombination is further restricted by epigenetic mechanisms to specific lymphocyte developmental stages. Such epigenetic mechanisms maintain antigen receptor genes in a closed conformation, so that the needed RAG-mediated double strand breaks that result in the formation of an antigen receptor take place at only the appropriate loci. For this reason, in the presence of RAG and despite large distances between gene segments, the *IgH* chain recombines to produce a functional antigen receptor in B-cells.

Even within a lineage, the different V, D, and J loci are rearranged in a specific manner. At the pro-B cell stage of B-cell development in the bone marrow D to J_H recombination precedes V_H to DJ_H recombination on the *IgH* locus. Due to the highly orchestrated regulation of RAG recombinase that results in the recombination steps

necessary for an antigen receptor, the different domains of the *IgH* locus must be accessible to the recombinase in a cell and stage-specific manner. Led by Fred Alt's laboratory, the "accessibility hypothesis" predicted the presence of accessibility control elements (ACEs) at the individual antigen receptor loci that promote the ordered genetic rearrangements (Yancopoulos and Alt, 1985). Using a combination of transgenic models and *in vitro* systems, both transcriptional enhancers and promoters were found to regulate the cell type-specific chromatin structure to determine the targeting of RAG recombinase. Later, the RAG nuclease-mediated cleavage of RSSs flanking the immunoglobulin gene segments was found to be developmentally regulated (Ji et al., 2010).

Though the mouse *IgH* locus consists of three main gene segments that must be accessible to RAG recombinase in a stage and site specific manner, the *IgH* locus also contains several *cis*-regulatory elements that govern epigenetic changes, including gene transcription, chromatin, and V(D)J regulation. This includes the intronic enhancer, Eμ, the 3' *IgH* regulatory region (*IgH* 3'RR), and transcriptional promoters throughout the V_H, D_H, and C_H gene segments. These *cis*-regulatory elements have a significant role in gene expression. They also serve as accessibility control elements (ACEs) to direct the common V(D)J recombinase for productive recombination events, and at the same, these *cis*-regulatory elements serve as a feedback mechanism to prevent further rearrangements from occurring.

Arguably the most critical *cis*-regulatory element within the *IgH* locus is the Eμ enhancer, located in the intron between J_H4 and the C_μ exons. It was the first transcriptional enhancer identified in mammalian cells and consists of a 220 base pair (bp) enhancer core (cEμ) and two flanking matrix attachment regions (MARs) (Gillies et

al., 1983; Jenuwein et al., 1997). Targeted deletion of these MARs showed they are not necessary for V(D)J recombination within the *IgH* locus (Sakai et al., 1999). However, deletion of cEμ in pro-B cells affected chromatin modifications, transcription and reduced both, D_H to J_H and V_H gene rearrangements, establishing this intronic enhancer as a foundational puzzle piece in B-cell antigen receptor development (Chakraborty et al., 2009). From this study and others, there was a correlation between recombinase accessibility and transcription that depended in large part on the presence or absence of the Eμ enhancer, demonstrating that enhancers are a means of regulating antigen receptor gene rearrangements, thereby allowing lymphocytes to respond to a diverse pool of individual antigens on their cell surface. In order to understand the critical importance of the *cis*-regulatory role of the Eμ enhancer in the *IgH* locus, this introductory overview will cover B-cell development and antibody structure, the organization of the *IgH* locus and functional rearrangement of V(D)J gene segments, and *cis*-regulatory features in the *IgH* locus. A clear understanding of this background will then enable transition remarks that illustrate how this thesis project builds upon previous experiments in novel ways.

B-Cell development and Antibody Structure

I. B-Cell Development: From Hematopoietic Stem Cells to Mature B-Cell

Predicated on F. Macfarlane Burnet's clonal selection theory (Burnet 1957), and supported with an array of experimental evidence (Cohn et al., 2007), the vertebrate immune system generates a large pool of B cells, each containing an antibody molecule with a unique specificity. In order for the "humoral" immune system to produce specific immunoglobulins targeted against a diverse array of pathogens, a variety of cell-surface

bound immunoglobulin species must be expressed on naïve B cells prior to antigen exposure. The multitude of immunoglobulins are produced in large part by the mechanisms that control V(D)J recombination, but also by the signals transduced from the pre-B cell receptor (BCR), which allows the developing progenitor B cell in the bone marrow to continue through the maturation steps as it exits and encounters antigen in the periphery. In brief, the major theme in B-cell development in the bone marrow is that *cis*-acting elements are necessary for successful gene rearrangement and the production of a protein, known as the B-cell receptor, which serves as a signal for the cell to enter the next stage of development.

Given the highlighted V(D)J rearrangement mechanism in the presence of RAG protein, the recombination events that take place between Ig gene segments are carefully regulated in a cell-specific manner. Known as allelic exclusion, most B cells express a single light (L) chain isotype, either Igκ or Igλ (isotype exclusion), and use only one of the two alleles of heavy (H) and light (L) chain genes. These parameters enable each B cell to express a single H_2L_2 combination on their cell surface. Different quality control steps ensure only a single H_2L_2 combination appears within the BCR. The immune system's ability to generate unique antibodies is dependent on the fact that prior to antigen stimulation, each naïve B-cell displays on its membrane several thousand identical copies of a single unique species of antibody, which serve as the BCR for that particular lymphocyte.

The journey of B cells from their earliest stage to their completed maturation form is characterized by both external and internal molecular events that are cell lineage specific. B cells are derived from a population of multifaceted pluripotent hematopoietic

stem cells (HSC) that reside in bone marrow niches optimal for their maintenance. These pluripotent HSCs have extensive self-renewal capacity and they are also able to regenerate all blood cell types throughout life by differentiating into progenitor cells with gradually restricted developmental potential. In the B-cell development pathway, an HSC first differentiates into multipotent progenitor cells (MPPs), which can produce both lymphoid and myeloid cells, but are no longer self-renewing stem cells. Multipotent progenitors express the tyrosine kinase cell-surface receptor FLT3, which binds to the membrane bound FLT3 ligand on stromal cells. Signaling through FLT3 is needed for differentiation to the next stage called the common lymphoid progenitor (CLP). While it used to be thought that the CLP gave rise to both the B-cell and T-cell lineages in culture, it remains unclear whether the CLP is a precursor for T cells. A preceding stage, known as the early lymphoid progenitor (ELP), has been identified that gives rise to T-cell precursors that migrate from the bone marrow to the thymus and also to the common lymphoid progenitor (Purizaca et al., 2012).

In order for precursor B-cells to be oriented towards their developmental pathway, a subpopulation of CLP cells differentiates into B220+ cells with the receptor for interleukin-7 (IL-7) present on their cell surface (Welinder et al., 2011). This receptor is induced by FLT3 signaling and PU.1 transcription, and is required for the survival and further differentiation from the CLP to the earliest B-lineage cell. The IL-7 cytokine is secreted by stromal cells and is required for both growth and maintenance of developing B-cells in mice. Two additional essential molecules are provided by the stromal cells. First, stem-cell factor (SCF), a membrane bound cytokine on stromal cells, interacts with the receptor tyrosine kinase Kit (CD117) and serves to stimulate the growth of

hematopoietic stem cells and the earliest precursor B-cells. Second, the chemokine CXCL12 is another essential factor produced constitutively by the stromal cells and is required for the early stages of B-cell development (Ratajczak et al., 2013). Therefore, there is strict requirement of stromal cells, for both the interleukins they secrete and the provided cell-cell contact during early B-cell development in the bone marrow.

Definitive B-cell commitment is tailored by the induction of several transcription factors. The current model is that IL-7 signaling promotes E2A expression, which then cooperates with PU.1 to induce the expression of early B-cell factor (EBF) (Medina and Singh, 2005). Together, E2A and EBF cooperate to drive the expression of proteins that determine the initial B-cell specific developmental step in the bone marrow, known as the pro-B cell stage (Figure 1A). This pair of transcription factors in the early pro-B cell stage also cooperate to induce the expression of the key proteins necessary for heavy-chain locus gene rearrangement, RAG1 and RAG2 of the V(D)J recombinase (Schatz and Swanson, 2011). In the absence of E2A or EBF, initial heavy-chain gene rearrangements fail to occur.

One other critical B-cell fate transcription factor is Pax5, also known as B-cell specific activator protein (BSAP), which is induced by E2A and EBF and suppresses alternate cell fates in early developing B lineage cells and activates B-cell specific genes such as CD19, Igα, and BLINK (Nutt et al., 1999). Experimental evidence demonstrates that germline deletion of Pax-5 in pro-B cells has no effect on D_H to J_H recombination, but blocks development further down the B-cell pathway, which indicates that Pax-5 is required for commitment to the B-cell lineage (Zhang et al., 2006). Two observations, one general and one Ig-specific, must be delivered here to understand the overall

development of B-cells in the bone marrow. First, even at this point of the pro-B cell stage and up to the immature B-cell stage, the progenitor B-cells remain in contact with stromal cells, particularly the reticular stromal cells in the trabecular spaces of the bone marrow. Then, with each stage maturation step, the developing B-cell migrates towards the central sinus of the marrow cavity until the immature B-cell journeys out of the bone marrow and into the peripheral lymphoid organs to become a mature B-cell. Second and more importantly, while the aforementioned transcription factors, cytokines and receptor-ligand interactions between common lymphoid progenitors and stromal cells give rise to the earliest B-lineage cell, the pro-B cell, the main theme from the pro-B cell stage to the immature B-cell stage is the fixed sequence rearrangement of one gene locus at a time (Figure 1A). For this reason, successful V(D)J rearrangement is necessary for the expression of a complete immunoglobulin chain as an integral component of the pre-B-cell receptor, and this initial heavy chain rearrangement is followed by a second rearrangement on the light-chain gene, which allows the immature B-cell to express a complete IgM antigen receptor at the cell surface where it is first tested for tolerance to self-antigens.

It was known from cell line studies that antigen receptor loci rearrangements first take place on the heavy chain, followed by the light chain (Figure 1A). In other words, D_H to J_H recombination takes place prior to V_H to DJ_H, and that this ordered rearrangement is followed by VJ_L recombination, resulting in a precise arrangement of steps for the antigen receptor loci (Alt et al., 1984). To map out the stages of B-cell development, it was an initial priority to define unique combinations of surface markers that could distinguish cells based on recombination status (Rajewsky 1996). This

resulted in the identification of the cluster of differentiation (CD) antigens, and their different combinations allowed for the separation of B-cell development into consecutive stages. Based off this mapping, as shown in Figure 1A, the order of B-cell development stages are as follows: pro-B cell, pre-BI cell, large pre-BII cell, small pre-BII cell, immature B cell, and mature resting B cell.

In light of this, the Ig loci of hematopoietic stem cells have the germline configuration, which is in contrast to the $B220^{+}$ $CD19^{-}$ $CD117^{+}$ early pro-B cells, where in the presence of RAG1 and RAG2, the *IgH* locus undergoes D_H to J_H joining on both alleles of the heavy chain locus. The diversity of the B-cell antigen-receptor repertoire is increased at the pro-B cell stage due to the presence of the enzyme terminal deoxynucleotidyl transferase (TdT). This enzyme adds non-templated nucleotides, known as N-nucleotides at the joints between rearranged gene segments. Also, later in the pro-B cell stage, a surrogate light chain (SLC) that contains λ5 and VpreB is expressed (Figure 1B). Expression of the SL chain is induced by E2A and EBF, and the SL chain is encoded with non-rearranged genes that are distinct from the antigen-receptor loci. The μ heavy chain assembles with the SL chain to form a "pre-BCR complex" at the cell surface (Figure 1B) (Karasuyama et al., 1990). A successful "in frame" V_H to DJ_H rearrangement leads to the production of intact μ heavy chains, which then stops any additional V_H to DJ_H rearrangement on the other allele, and the cell is now a pre-BI cell ($B220^{+}$ $CD19^{+}$ $CD117^{+}$). A second rearrangement proceeds on one allele between a V_H gene segment and the rearranged DJ_H sequence. If both V(D)J recombination products are out of frame on the second allele, the pro-B cell is then eliminated by apoptosis and

no μ heavy chains are produced. It is estimated that 45% of pro-B cells are lost at this stage (Fang et al., 1996).

On these pre-BI cells, the pre-BCR associates with Igα and Igβ on the cell surface, which are the signaling molecules for this receptor (Figure 1B) (Clark et al., 2005; Martensson et al., 2007). The expression of the pre-BCR complex has been shown to have importance on two levels. First, it blocks additional *IgH* chain recombination by lowering both RAG expression (Grawunder et al., 1995) and V_H transcription to reduce target accessibility (Schlissel et al., 1994). This allows for a single rearranged V(D)J antigen receptor to be expressed in the pre-BCR. If successful rearrangements happened at each Ig heavy chain allele, this would result in a B-cell producing two receptors of different antigen specificities. Allelic exclusion prevents this so that only one of two alleles of a gene is expressed in a diploid cell (Melchers et al., 1999). The pre-BCR continues to be expressed until the next maturation step, known as the large cycling pre-BII cell stage ($B220^+$ $CD19^+$ $CD25^+$). Several rounds of cell division take place in this stage, and it is estimated that the population of cells with in-frame gene segments expands by 30 to 60 fold prior to the next stage (Hesslein et al., 2001). RAG gene expression declines in large pre-BII cells, and after several divisions the cells become smaller and stop dividing. As these cells exit the cell cycle and become small pre-BII cells, the SL chain is no longer detected, and second importance of the pre-BCR is observed, in that it initiates immunoglobulin light chain recombination. RAG gene expression is turned up once more, and now the immunoglobulin light chain genes undergo recombination. Complete synthesis of either the κ or λ light chain leads to the assembly of the B-cell receptor (BCR). Similar to the heavy chain locus, the light chain

locus is also governed by the same allelic exclusion mechanism so that the lymphocyte expresses only one type of antigen receptor (Loffert et al., 1996). Most B cells also show isotype exclusion, in that they express either the κ or λ light chain, but not both (Ma et al., 1992). Of particular note with immunoglobulin light chains is that they lack D-gene segments, and therefore rearrangement only takes place by V to J joining.

The BCR associates with Igαβ and is expressed on the cell membrane as an anchored IgM antibody. This cell is now an immature B cell ($CD93^+$) and RAG expression is again decreased. Immature B cells in the bone marrow now contact autoantigens via the BCR and are assessed for autoreactivity. Such autoreactive BCRs may be replaced by non-autoreactive BCRs in a mechanism known as receptor editing, wherein a new light chain is generated through the expression of RAG. After receptor editing, the IgM^+ immature B cells also express IgD on their cell surface and migrate to the periphery and wait for antigen exposure (Figure 1A). In peripheral lymphoid organs (spleen, lymph nodes), immature B cells further develop into resting naïve mature B cells ($CD93^-$) that co-express both IgM and IgD. The *IgH* locus in resting B cells exists somewhat in a closed chromatin state enriched in repressive histone modifications (Chowdhury et al., 2008; Jeevan-Raj et al., 2011). In contrast, active histone marks are observed throughout the V(D)J region, including Eμ and the *IgH* 3'RR and the open chromatin footprint remains from the early gene recombination events in pro-B cells (Daniel et al., 2010; Wang et al., 2009). Transcription still takes place in resting B cells throughout the *IgH* locus, starting at the V_H promoter and going through the intronic switch (Sμ) region and Cμ/Cδ exons. This encodes the BCR that consists of the Igμ and Igδ heavy chain genes.

Once resting naive mature B cells encounter antigen in a secondary lymphoid organ, they proliferate and then differentiate into germinal center B cells, which give rise to long-lived plasma cells and memory B cells. Upon this event of antigen stimulation, B-cells undergo class switch recombination (CSR) to produce antibodies with the same variable region, but different constant region, allowing for different effector functions. The central mechanism is one by which the intervening sequences between two switch (S) regions that are upstream of the constant (C) region are deleted in response to specific cytokine stimulation from helper T cells. This allows the assembled V_H gene to be next to a new C-region that has a different effector function, but has the same antigen specificity. This process is initiated by activation-induced cytidine deaminase (AID). AID targets the S region and converts cytosine to uracil, and the repairs of the guanine:uracil lesion lead to double-strand breaks and class switch recombination (Muramatsu et al., 2000).

Another process, somatic hypermutation, also takes place in the germinal center of activated B cells. It modifies the gene encoding the variable region of immunoglobulin heavy and light chains, and generates B cells with higher affinity antibody. Somatic hypermutation contributes to the affinity maturation in the antibody response by inserting point mutations in Ig V(D)J DNA, which results in high affinity antibodies that are selected for antigen. It has also been suggested as a mechanism to assist the host in finding evolving pathogens (Long and Lipsky, 2006). Only a small fraction of the B cells express a BCR capable of binding to any particular antigen. When these B cells bind their antigen, they become activated to proliferate and mature into antibody secreting plasma cells, which manufacture large amounts of antibody specific

for the activating antigen. In order for any antibody to effectively bind antigen with high affinity and avidity, the immunoglobulin structure is critical to ensure a diverse response.

II. Antibody Structure

The ability of the immune system to generate an antigen specific antibody depends on the fact that, prior to antigen exposure, millions of naïve resting B cells circulate and each of these cells displays on its membrane several thousand identical copies of a single unique species of antibody that serve as B-cell receptors (BCRs) for that lymphocyte. Though unique, each antibody shares many structural characteristics to allow for this intense diversity.

Each antibody molecule is comprised of two identical heavy (H) and two identical (L) chains (Figure 2). These chains are held together by disulfide bonds to form a symmetric Y-shaped tetramer. Due to sequence similarity, each heavy and light chain can be sub-divided into two general regions, the variable region and the constant region. The variable region contains the antigen binding domain and is therefore highly polymorphic. Yet, the specificity of each antibody is determined by the combination of the heavy and light chain variable regions. This variable region can be further divided into four framework regions (FR) and three complementarity determining regions (CDR) (Kabat et al., 1991). The CDRs are hypervariable and the FRs forms a beta-sheet structure that allows for a scaffold to hold CDRs in positions to contact antigens. Interestingly, there are fewer D_H and J_H gene segments than V_H genes; however, the recombined DJ_H structure contributes significantly to the diversity of the variable part of the immunoglobulin structure because it encodes most of CDR3 (Atkinson et al., 1994).

While the specificity of each antibody is determined by the variable region, the constant region determines the mechanisms used to destroy antigens. Mouse antibodies are divided into five different classes based on their structure and function: IgM, IgG, IgA, IgE, and IgD. Intense diversity at the variable region and a variety of mechanisms guided by the constant region is needed at this stage in order to combat a diverse universe of pathogens. The diversity of these BCRs prior to antigen exposure is dependent on the structure and function of the immunoglobulin locus. In turn, the ability for a B-cell to express a functional BCR is directly dependent on the organization of the *IgH* regulatory domain and the *cis*-rearrangements of gene segments within this domain.

IgH Locus Organization and Functional Rearrangement of V(D)J Gene Segments

I. Structure of the *IgH* Locus

The loci encoding *IgH* is the largest multigene domain known in the mammalian genome and its structure is critical for one of the most complicated genetic functions in the nucleus, V(D)J recombination (Bolland et al., 2009). The murine germline *IgH* locus may be subdivided into four major gene clusters, spread over nearly 3 Mb on the long arm of chromosome 12 (Zhou et al., 2002, Retter et al., 2007). As shown in Figure 3, these regions include the variable (V_H), diversity (D_H), and joining (J_H) gene segments, all upstream of several constant region exons. The V_H gene segments in mouse contains nearly 110 functional genes and 85 pseudogenes that are classified on sequence similarity into 16 V_H gene families and are distributed over 2.5 Mb in the 5' part of the locus (Johnston et al., 2006). Three broad domains characterize the V_H gene cluster as follows: (1) an interleukin-7 (IL-7) regulated domain that is about 5 Mb from the telomere and

consists of the largest V_H family, the V_HJ558 family, (2) an intermediate domain that consists of the V_HQ52 family, and (3) the proximal 3' domain that contains the V_H7183 gene family (Johnston et al., 2006). Each functional V_H gene has a lymphoid-specific promoter, followed by a leader sequence, leader intron, V_H coding sequences, and recombination signal sequence (RSS). These RSSs are the recognition sequences for the V(D)J recombinase proteins RAG1 and RAG2.

Separated by a distance of 90 kilobases (kb) from the most proximal 3' V_H gene (7183.2.3) is the D_H gene domain, which has a total length of ~52 kb and contains 10-15 (Figure 3). These gene segments can be grouped into four families based on sequence homology: DFL16, DSP2, DST4, and DQ52. Each functional D_H gene is flanked by one RSS on each end. DFL16.1 is the most 5' D_H gene segment, which is then followed by the DSP gene segments that are repeated in 4.7 kb repeats. DQ52 is the most proximal D-gene and is only separated by 700 bp from the next gene-segment of the *IgH* locus, the J_H gene cluster. There are 4 functional J_H gene segments that are all located within a 1.3 kb region. Each J_H gene segments is flanked by an RSS at the 5' end. Less than 300 bp from the last J_H gene is the Eμ intronic enhancer, a 1 kb Xba I fragment between J_H4 and the first Cμ exon.

Downstream of the Eμ intronic enhancer are eight constant heavy (C_H) regions that have between 4-6 exons, which include Cμ, Cδ, Cγ3, Cγ1, Cγ2b, Cγ2a, Cε, and Cα (Figure 3). Each of these gene segments encodes three (α, δ and γ) or four (μ and ε) domains (CH1-3 or CH1-4) and are separated by large introns. Each C_H region encodes one of the different immunoglobulin isotypes (IgM, IgD, IgG, IgA, and IgE) (Dudley et al., 2005). In the case of Igα, Igδ, and Igγ chains, a proline-rich hinge region is present

between the CH1 and CH2 domains, and this region allows for structural flexibility that assists in the binding of antibody to an antigen. Except for Cδ, each constant region gene has its own intronic promoter. Transcripts derived from these promoters are "sterile transcripts" and are not translated into proteins, but correlate with CSR.

II. Functional Rearrangement of V(D)J Gene Segments

An understanding of the *IgH* locus structure is foundational to exploring how the different gene clusters are imprecisely shuffled and rearranged to achieve the vast repertoire of custom-made antibodies for adaptive immunity. Once this groundwork is laid, an overview of the important *cis*-regulatory features of the *IgH* locus will be outlined in order to present the main design and thesis of this project. For, it is the precise position of these *cis*-regulatory elements that allow for the rearrangement of separate these V_H, D_H, and J_H gene clusters.

Diversity lies at the core of V(D)J recombination and is consequently the cornerstone mechanism that is responsible for the variety of immunoglobulin receptors. Such diversity comes from two sources. First, combinatorial diversity is due to the numerous combinations of V, D, and J segments that may be arranged for different receptor specificity. Second, junctional diversity stems from the imprecise joining of the V, D, and J gene segments.

In order for recombination to take place between different V, D, and J gene segments, an RSS must be present at each unrearranged gene segment, and RAG recombinase must be recruited to the loci specified for rearrangement (Figure 4A). These evolutionary conserved sequences represent direct targets for RAG nuclease. Each RSS

consists of a highly conserved 7-bp sequence, known as the consensus heptamer (5'-CACAGTG), and an AT-rich 9-bp nonamer consensus sequence (5'-ACAAAAACC). These sequences are separated by either 12 or 23-bp spacer sequence that is less conserved. As a result, spacer length defines the two types of RSSs, termed 12-RSS and 23-RSS. Efficient recombination happens only between 12-RSS and 23-RSS, a restriction known as the 12/23 rule. Also required for the cleavage reaction *in vitro* is Mg^{+2}, while Mn^{+2} yields RAG-mediated nicking of a single RSS (van Gent et al., 1996).

Following RAG binding, V(D)J recombination is initiated with the assistance of high mobility group proteins (HMG1/2), and synapsis takes place between segments that are flanked with dissymmetric RSSs (Figure 4A). *In vitro* biochemical studies of RAG-mediated RSS cleavage have provided evidence of a two-step 'capture' model of RSS synapsis involving RAG binding to a first RSS prior to the capture of a fit partner (Jones et al., 2002; Mundy et al., 2002). It was even proposed that the detection of initial cleavage mostly occurred at 12-RSSs *in vivo*, which resulted in a so called '12RSS binding first' model (Curry et al., 2005). However, other studies have not shown such preferential RAG binding to either RSS cluster (Ji et al., 2010).

After the gene segments are brought together, RAG nuclease cleaves the RSSs by nicking a single strand of DNA specifically at the border between the heptamer of RSS and the coding segment, which results in a double-strand break (Schlissel et al., 1993). The 3'-OH group at the nick site of the coding segment is then covalently linked to the opposing phosphodiester bond on the antiparallel strand by a transesterification reaction that results in a hairpin structure, which contains the nearby coding DNA, and a blunt signal end (Roth et al., 1993). The broken ends are then processed and joined together

with the assistance of ubiquitously expressed DNA repair factors, including members of the non-homologous end joining pathway (DNA-dependent protein kinase, Ku, Artemis, DNA ligase IV, Cernunnos/XLF, Xrcc4 proteins, along with histone H2AX and the Mre11/Rad50/Nbs1 complex) (Lieber, 2010; Boboila et al., 2012). The result of these interactions are a signal joint (SJ) and a coding joint (CJ) product that contains the exons and forms the antigen-binding region of the receptor. The SJ structure is a back-to-back fusion of RSS heptamers and the CJ structure displays sequence variability at DNA junctions because of the nucleotide deletion and non-templated nucleotide addition. The V(D)J rearranged gene segment maintains a correct reading frame in approximately one-third of instances, and out-of-frame reading frames of CJ to account for the remaining two-thirds. The one productive rearrangement then enables development of a potentially functional immunoglobulin at the pre-BCR stage of B-cell development in the bone marrow.

Three levels of control have been proposed to localize the activity of V(D)J recombinase. As previously mentioned, V(D)J recombination is specifically coordinated to the cell-lineage and developmental stage of the differentiating B-cell in the bone marrow. There is a hierarchy for *IgH* rearrangements in pro-B cells, in that it starts with D_H to J_H rearrangement before the completion of the rearranged allele with the additional V gene segment (Subrahmanyam et al., 2012) (Figure 4B). Moreover, *IgH* genes rearrange first, which is then followed by *IgL* genes. In addition, another level of control that constrains the activity of V(D)J recombinase is the phenomenon of allelic exclusion. In this regard, Burnet's clonal selection theory appears again, as the antigen receptor chain is only encoded by one of two alleles from the heavy and light chains, and this

ensures the specificity of the immune response depending on the antigen selection of discrete B cells restricted to a homogenous set of Ig receptors. When a functionally relevant rearrangement that encodes for *IgH* and can associate with the pre-BCR, allelic exclusion results where through feedback inhibition, another rearrangement at the corresponding locus is prevented. The "accessibility hypothesis" was originally formulated out of this context in an attempt to interpret cell-lineage and developmental stage specificities of V(D)J recombination, and in particular the accessibility hypothesis sought to explain the how the feedback inhibition down-modulates the accessibility of the remaining allele (Alt et al., 1984).

The central prediction of the accessibility hypothesis is the existence of "accessibility control elements" (ACEs) at the individual antigen receptor genes that are required for the complete V(D)J recombination mechanism to proceed. Early transgenic mouse knockout studies identified both promoters and enhancers on Ig loci as ACEs that act to make the locus accessible to V(D)J recombinase (Ferrier et al., 1990; Bassing et al., 2002). However, mutations in germline enhancers led to inhibition of V(D)J recombination locus-wide, whereas mutations in germline promoters demonstrated localized inaccessibility of gene segments for recombination (Whitehurst et al., 1999). Yet, mutations in each case resulted in epigenetic chromatin changes (Whitehurst et al., 2000), and targeted gene mutation of chromatin modifying factors that were known to bind to different ACEs of the *IgH* locus were shown to affect V(D)J recombination (Lazorchak et al., 2006; Johnson et al., 2008). This finding was transitional for the accessibility hypothesis as it pointed to chromatin structure as one of the determining factors for V(D)J recombinase.

Indeed, using an *in vitro* system to investigate RAG mediated cleavage of RSSs flanking Ig gene segments in cell nuclei, Schlissel's group demonstrated that chromatin structure is the machinery that targets V(D)J recombinase (Stanhope-Baker et al., 1996). They found a temporal ordering of gene segment rearrangement that is in line with cell type specificity to be dictated by the chromatin structure, which renders the accessibility of RSSs to RAG cleavage. This was followed up with several chromatin immunoprecipitation experiments using antibodies raised against activating or suppressing histone modifications. Many groups demonstrated the accessibility of the immunoglobulin loci to be dependent on epigenetic histone marks associated with gene activation, and the corollary to hold true, inaccessible immunoglobulin loci were marked with histone marks associated with gene silencing (Chowdhury and Sen, 2001; Morshead et al., 2003; Espinoza et al., 2005; Chakraborty et al., 2009). In the case of the former, this includes histone acetylation for H3 and H4, and di/tri-methylation of lysine-4 on histone H3 (H3K4me2/me3), while in the case of the latter, this includes di/tri-methylation of lysine-9 or lysine-27 on histone H3 (H3K9me2, H3K27Me3). Further, a hierarchal establishment of locus-specific chromosome accessibility was established through the identification of discrete domains of RAG nuclease recruitment. Such "hotspot" domains were associated with H3K4me2/me3, and the activating marks were significantly reduced by enhancer mutation (Spicuglia et al., 2002; Chakraborty et al., 2009) (Figure 6).

Several other genetic manipulation experiments have been executed to intentionally wield the accessibility of V(D)J recombination at individual gene segments. In one instance, the histone methyl transferase enzyme G9a, which mediates H3K9

methylation, was shown to override the ACEs function and diminish V(D)J recombination of chromosomal gene segments in the TCR locus (Osipovich et al., 2004). This included several histone alterations and showed how histone methyltransferases can revise the local chromatin environment and impair ACEs in a tissue and stage-specific manner during lymphocyte development. Thus, the experimental weight pointed towards chromatin modifications as the major key that explains the accessibility hypothesis and the ACEs, as lineage specific RAG accessible and inaccessible immunoglobulin loci were discovered. Nonetheless, these revisions were not isolated to the chromatin environment alone. When a transcription terminator was introduced to block transcriptional elongation, this also suppressed chromatin remodeling and reduced recombination (Abarrategui et al., 2006). As a result, the picture that emerged was that epigenetic chromatin modifications introduced changes at the level of DNA, which affected both transcription and the ability of chromatin remodeling complexes to remodel nucleosomes positioned at RSSs to either increase RSS accessibility to RAG cleavage or recruit RAG itself (Figure 6).

In order to elucidate the mechanism of RAG nuclease binding to RSSs of antigen receptor gene segments, it is first helpful to understand some background and the functional significance of the RAG protein. RAG1 and RAG2 are two evolutionary conserved genes that are closely linked on chromosome 11 in humans and chromosome 2 in mice. RAG1 consists of a total of 1040 amino acids and has a core region that is essential for all *in vivo* and *in vitro* activities (amino acids 384-1008) (Sadofsky et al., 1993; Silver et al., 1993). RAG2 has 527 amino acids is comprised of a non-canonical plant homeodomain finger where, in several other proteins who function in epigenetic

regulation, binds to H3K4me2 or H3K4me3. It has been demonstrated that this same domain directs the binding of RAG2 to H3K4me2 and H3K4me3 (Liu et al., 2007). The functional significance of this binding by RAG has been demonstrated on three levels. First, mutations in this plant homeodomain removed H3K4me3 recognition, which then resulted in significantly lower amounts of V(D)J recombination (Liu et al., 2007). Second, the amount of H3K4me3 dictated the level of RAG DNA binding and recombination. Third, RAG2 recognition of H3K4me3 stimulates the catalytic activity of the RAG nuclease, and consequently is likely to stabilize the newly excised recombination ends within the post-RAG cleavage complex that contains the non-homologous end joining repair machinery (Grundy et al., 2010; Shimazaki et al., 2009; Wang et al., 2012).

This sets the stage to understand the binding of the RAG nuclease to RSS-containing sites of antigen receptor gene segments. Led by David Schatz's group, they executed several significant experiments using transgenic mice that expressed a RAG1 protein and, while on the one hand binds DNA normally and interacts with RAG2, contained a mutated site in the active domain that resulted in no catalytic activity (Ji et al., 2010). This mutation was at three residues (R391A, R393A, and R402A) and known to make critical contacts with the RSS nonamer. It was shown that in the mutated RAG1 setting, though RAG1 is expressed, lymphocyte development does not take place presumably due to no V(D)J recombination. It was further established that RAG protein DNA binding takes place in a focal manner in minimal regions that are elevated in the active histone mark H3K4Me3 and have a RSS. Formations of these local RAG-bound regions are referred to as "recombination centers," and occur in a developmental stage

and lineage-specific manner. The goal of these recombination centers is to coordinate V(D)J recombination by providing discrete sites within which antigen receptor gene segments are captured for recombination (Jaeger et al., 2013). Strikingly, it was found that on the TCR locus, it is the enhancers that control RAG1 binding globally in the DJ gene segment, and that promoters direct RAG1 locally in a manner that recapitulates the V(D)J gene rearrangement function of these accessible control elements as defined in earlier studies (Ji et al., 2010).

To close this section, two hallmark features have emerged from recombination centers. First, they are specialized site of high local RAG concentration associated with high active chromatin marks and second, where they exist, there is an array of multiple, closely spaced RSSs. It remains to be determined whether the number and/or spacing of the RSSs are important for stable recruitment of the RAG proteins. Even recent experimental evidence from recombined DJ_H alleles on the *IgH* locus in pro-B cells indicates that these DJ_H junctions bind RAG proteins and are epigenetically marked to permit the DJ_H-5' RSSs to start the second step of *IgH* gene assembly (Subrahmanyam et al., 2012). Therefore, while the accessibility hypothesis initially opened the case to explain allelic exclusion, in terms of the accessibility of V(D)J recombinase to ACEs through the opening or closing of chromatin structure in a cell-lineage and developmental stage specific manner, it was the "recombination center" model that has further explained the strict requirement for high local concentration of RAG tethered to an array of RSSs as necessary for each ordered V(D)J recombination step. Together, each model, the accessibility hypothesis and recombination center, work in concert to explain the different levels of regulation for the functional V(D)J recombination mechanisms. It is

precisely the accessibility of both an antigen receptor gene locus, and its ACEs, to the RAG recombinase that determines the specific selection of an antigen-receptor gene that will eventually recombine. Consequently, these ACEs, namely enhancers, along with promoters and locus control regions, are then considered to be the "controllers" of RAG binding and the eventual recombination of genes. Specifically, ACEs act in concert to open the chromatin structure and free the RSSs for RAG-binding to take place. Such *cis*-acting ACEs are then central to the core features of both the accessibility hypothesis and the posited recombination centers.

In light of this, it is critical to give an overview of the *cis*-regulatory elements of the *IgH* locus on mouse chromosome 12. From such an overview, an explanation will follow that communicates the significant role of the Eμ intronic enhancer and why it was selected for this thesis as an enhancer to study both *cis*-regulation as it relates to recombination, and *cis*-regulation to other, unknown areas of the *IgH* locus.

Significant cis-Regulatory Features in the IgH locus

I. Overview of *cis*-Regulatory Elements

From the earliest stages of B-cell development in the bone marrow to the eventual maturation of an immature IgM B-cell, the differentiation into the tissue and stage-specific cell types requires spatiotemporal patterns of gene expression and chromatin modifications that are regulated by *cis*-regulatory features in the non-coding DNA sequence. These *cis*-regulatory elements control the nearby expression of genes and recruit multiple nuclear factors in order to ensure precise *IgH* locus regulation through directed physical interactions and potential functional consequences. Prior to an

examination of the *cis*-acting genetic features on the *IgH* locus, an overview of *cis*-regulatory elements initiates this section in order to better understand those elements on the *IgH* locus and the central priority of this thesis.

In general, *cis*-regulatory elements include enhancers, promoters, chromatin insulators, silencers, and locus control regions. Chromatin insulators serve to prevent the spread of inactive chromatin domains into the active gene cluster (Gaszner et al., 2006), and silencers are DNA sequences located typically upstream of the target gene and represses transcription. Yet, the formation, organization and maintenance of many open chromatin domains is the role of the locus control region (LCR) (Li et al., 1999). Similar to insulators, an LCR can overcome the positional effect when linked to a new gene that is inserted into a eukaryotic chromosome and enhance the expression of that gene to physiological levels in a tissue-specific and copy number dependent manner. Distinct from insulators, an LCR also stimulates the expression of genes contained within its functional domain.

Unlike other *cis*-regulatory elements, the location and composition of LCRs relative to their cognate genes is not uniform. The discovery of the position, integration-site independent, and copy-number dependent LCR in the β-globin locus to drive the high expression of genes from transgenic mouse studies led to deletion studies of regions upstream of the gene (Grosveld et al., 1987), which resulted in a closed chromatin conformation spanning the whole locus and suppression of gene expression (Forrester et al. 1990). This data suggested that the deleted DNA segment contained an indispensable *cis*-acting regulatory element required for beta –globin expression in vivo. Other experiments centered on similar model systems demonstrated long-range gene activation

via LCR-mediated targeting and extensive spreading of core histone acetylation (Ho et al., 2002). The LCR was therefore assumed to control chromatin structure and form active chromatin domains for gene activation. While small deletions within individual LCR hypersensitive sites in transgenic human loci have been shown to result in gene and developmental-specific effects on LCR activity in of embryonic and fetal globin genes (Navas et al., 2003), no similar results have been obtained from endogenous loci and therefore the developmental specificity appears to be encoded elsewhere with the β-globin LCR. Evidence to date indicates that the endogenous mouse β-globin locus LCR is only present for enhancer-based activation of gene expression and the controllers of chromatin structure are in other *cis*-acting elements as LCR deleted alleles have regular levels of histone acetylation but significant reduction in β-globin transcription (Schubeler et al., 2001). To summarize, the characterization of LCRs points to two significance features: (1) the regulation of gene expression by LCRs are developmental and cell lineage-specific, and (2) the reliance on both gene-proximal *cis*-acting elements and long-range interactions of *cis*-regulatory elements may drive dynamic chromatin alterations to bring about the necessary patterns of gene expression for tissue specific cellular differentiation.

While LCRs are *cis*-acting elements that organize gene clusters into an active chromatin domain for the purposes of enhancing transcription of genes in a tissue specific manner, it is the specific role of other *cis*-regulating elements, namely enhancers that coordinate with promoters to activate transcription. One of the primary functions of an enhancer is to recruit DNA-binding proteins to the gene locus and provide diverse transcriptional instructions in order to regulate the level of gene expression necessary for

biological function. Consequently, mutations in enhancers result in abnormal gene expression and may be the source for various disorders (Cho, 2012). Enhancers are usually occupied by clusters of transcription factors that are free of nucleosomes and thus contain a distinguishing hypersensitive character; that is, one distinguishing feature of enhancers is their sensitivity to small amounts of DNase I as the chromatin is digested due to the open structural conformation.

Enhancers are critically important *cis*-regulated gene expression controllers because it has long been experimentally suggested from transfection assays that "enhancers can activate transcription independent of their location, distance, or orientation with respect to the promoters of genes" (Ong et al., 2011; Banerji et al., 1981). Such a view is derived from positioning enhancers 5' or 3' to reporter genes and observing enhancer-mediated transcription. These *cis*-regulatory enhancers are distinct from promoters, as promoters are known to act on specific genes in a close distance to the transcription initiation site. Enhancers are generally situated at considerable distances from the 5' and 3' gene transcription sites, and from the historical view, thought to function independently of orientation and distance to the transcription initiation site (Bulger et al., 2010).

The current estimate for the number of enhancers in the mammalian genome is about 1 million (Heintzman et al., 2009; Vaquerizas et al., 2009). If the average size of enhancer is assumed to be approximately 100 bp in length, this would translate to a total combined length of 10^8 bp for all enhancers, which is 3.5% of the haploid mouse genome. In the human genome, by this calculation enhancers make up 3% of the DNA, and this number exceeds the coding region size (1.5%) (IHGSC, 2001). Therefore, the

critical biological role of enhancers is established from the outset. Coupled to this, the identification of new regulatory elements remains a significant task to understanding mammalian genome regulation, especially as it relates to the complex immunoglobulin locus.

One of the themes of enhancer and promoter communication is the particular *cis* relationship highlighted in this section. Particularly, the central dogma of biology is that expression of genes is regulated by DNA sequences that act in *cis* on the same chromosome and respond to RNA and proteins that are encoded by genes acting in *trans* on different chromosomes. This is supported by studies from the beta-globin locus where enhancers and LCRs augment the activity of promoters by interactions that consist of looping, which is the bringing together of distant DNA sequences into such close proximity that a loop forms (Chakalova et al., 2005). Other experimental observations also point to *trans*-regulation by enhancers. For example, in *Drosophila* where homologous chromosomes are paired in somatic cells, enhancers are associated with one allele can activate the promoter of a second allele in a process coined as "transvection" (Duncan 2002). Even transfection assays of artificial gene constructs demonstrated that an enhancer can activate a promoter that is topologically unlinked, but is associated with either a protein bridge or located on a separate, but interlocked plasmid (Mueller-Storm et al., 1989).

II. Overview of *cis*-Regulatory Elements: Chromosomal Interactions

In the mammalian setting, physical and functional evidence has established non-allelic interactions between chromosomes. An LCR in the interferon-γ locus was shown

to associate with the interleukin-4 locus on a different chromosome in naïve T cells that are committed to differentiate into cells expressing only one of two cytokines in a mono-allelic manner (Spilianakis et al., 2005). Similar non-allelic interactions between chromosomes 7 and 11 in mice have been shown in the imprinting control region of the *Igf2/H19* locus and the *Wsb1/Nf1* gene (Ling et al., 2006). Lomvardas and colleagues have experimental evidence where a single enhancer of a murine odorant receptor (OR) gene cluster is able to interact with multiple OR gene promoters on different chromosomes, but associates with only one OR gene in any given olfactory neuron (Lomvardas et al., 2006). This study suggested a mechanism that allows olfactory sensory neurons to choose randomly and express only one out of more than 1000 OR genes.

There is also data in the human setting that proposes inter-chromosomal associations can form rapidly in response to extracellular cues. For example, estrogen-inducible genes are rapidly expressed when treated with 17β-estradiol (E2). However, most estrogen receptor (ER-α) binding sites are intergenic and distal from E2-inducible genes, which suggests they form long-range looping associations (Lin et al., 2007). To explore this possibility, Hu and colleagues developed a novel variant of 3C technology termed ‘deconvolution of DNA interactions by DSL (3D) (Hu et al., 2008). Using this approach, they identified several intra-chromosomal associations between the E2-regulated trefoil factor 1 (TFF1) gene on human chromosome 21 and other ER-α bound loci on the same chromosome. In addition, they identified an inter-chromosomal association between TFF1 and an E2-regulated gene, which is known to be regulated by estrogen in breast cancer protein (GREB1), on chromosome 2. In untreated cells these

associations were absent but formed rapidly (within 15 min) on treatment with E2. Remarkably, these associations were paralleled by re-localization of the entire nuclear territories of chromosome 21 and chromosome 2, which became intimately associated.

In order to explain such *cis and trans* regulatory interactions between gene elements in different locations, a three-dimensional view of the nucleus is helpful. For, it is not simply the organization of genetic elements within an individual locus that are critical for tissue-specific regulation, but the organization of the genome within the nuclear space has implications for orchestrating gene expression. Chromatin surfaces again in this regards, as an accumulation of experimental evidence demonstrates that chromatin is folded into loops that bring regulatory elements into close contact with each other. In this regard, it has been well documented that heterochromatin and euchromatin segregate within the nucleus, forming chromatin neighborhoods with similar properties (Dillon, 2004). It has also been shown that gene-rich chromosomes are located preferentially towards the center of the nucleus, a feature conserved through evolution (Neusser et al., 2007). Data from chromosome conformation capture (3C) studies has allowed analysis of chromosome folding and has revealed how promoters communicate with distal regulatory elements. In mammalian cells, 3C was first used to investigate promoter–enhancer communication at the β-globin locus (Tolhuis et al., 2002). These studies revealed that, through DNA looping, hypersensitive sites within the LCR come into close physical proximity with the active globin genes situated 40–60 kb away, forming a structure called the active chromatin hub (ACH). Additional distal hypersensitive sites were found in close association with the LCR and active globin genes. These interactions were mediated by CTCF, which is an 11-zinc-finger nuclear

protein that has roles in transcriptional insulation and chromatin boundary formation (Phillips et al., 2009). CTCF was found to be bound to the 5' and 3' ends of the locus (Splinter et al., 2006). The intervening sequences and olfactory receptor genes were looped out of this complex. It has been suggested that the clustering of regulatory elements within the ACH increases the local concentration of transcription factors and maintains active chromatin domains, facilitating high levels of transcription (Palstra et al., 2008). The presence of long range loops have been observed in other loci, including those of antigen receptors, and cytokines in lymphocytes (Wuerffel et al., 2007; Sekimata et al., 2009).

III. Overview of *cis*-Regulatory Elements: Enhancers and Gene Regulation

Given the above background, it is of particular interest for the purposes of this thesis to study an enhancer's ability to affect gene regulation, in terms of chromatin, transcription and recombination, with respect to distance. Such a study will result in a more complete understanding of gene regulation. While looping studies, using either ligation based methods to measure proximity between regions (3C, 4C) or direct microscopic visualization (3D-FISH), are helpful to analyze potential regions of regulation and transcription hubs, these approaches do not directly address the *cis* distance of an enhancers effect on gene regulation. In the case of the locus control region (LCR) of the murine β-globin, looping results have left open the search for additional *cis*-acting elements distant from target genes and then follow-up studies for how these *cis*-sequences find the target gene. In prokaryotic cells, enhancers have been reported to control gene regulation up to several kb (Bondarenko et al., 2003) and enhancer

communication for the mouse *Sonic hedgehog* (*Shh*) gene has been reported greater than 400 kb from the transcription start site (Sagai et al., 2009). Yet, the role of chromatin was not explored. In another setting, the human insulin (INS) locus in pancreatic islets on chromosome 11, which is also home to *IGF2*, is flanked by binding sites for CTCF and is part of an extensive, active chromatin domain that has marks for H3K4me2, H3ac, and H4ac (Mutskov and Felsenfeld. 2009). Low levels of the inactive chromatin histone H3K27me3 are also detected across this domain. Another gene, *SYT8*, is located 300 kb away on chromosome 11 and also has CTCF binding sites. Homologues for *SYT8*, *Syt7* and *Syt9*, have been demonstrated to regulate insulin secretion in mouse islets (Gustavsson et al., 2008). In the presence of glucose, both *SYT8* and *INS* expression was increased and depletion of CTCF resulted in lower levels of *SYT8* expression, but *INS* expression was not affected (Xu et al., 2011). This data was corroborated with 3C measurements that showed an interaction between the two sites. These results established the existence of a regulatory network where a *cis*-acting element, a promoter located 300 kb away from its target gene, is able to regulate expression in the presence of another gene, and this expression is strengthened in the presence of glucose. Understanding why the chromatin structure is open at *INS* remains to be investigated, but this study established a role for *cis*-regulatory elements to regulate gene expression and added support to studying whether the particular location of a *cis*-acting element is critical for its function.

From the preceding overview, two major themes emerge for *cis*-regulatory features and their ability to alter the expression of nearby genes. First, in regards to enhancers, although such mechanisms are well documented from plasmid transfection

assays and in textbooks (Krebs et al., 2012), it is unlikely to be explanatory *in vivo* because it fails to account for: (1) the cellular environment of genes, which includes the aforementioned *cis*-regulatory elements, and chromatin, nucleosomes, and transcription binding proteins, (2) the preferential bias of enhancers to initiate transcription from one set of promoters and not others, (3) the unknown effect of moving enhancers to different positions within a cellular gene, and (4) the developmentally regulated activity of enhancers that is likely to change based off cell differentiation status or time. As a result, it is important to investigate the *cis*-activating effects of an enhancer *in vivo*, and to determine any changes in gene expression and chromatin as a result of moving the enhancer to different positions within a cellular gene. Such an investigative study will either confirm the current model of enhancers established in *in vitro* systems, where an enhancer's effect is location independent, or from the same *in-vivo* setting, incorporate a viewpoint that takes into consideration the presence of both active and silent chromatin histone markers and the associated promoters of these regions.

Second, the preceding background on *cis*-regulating elements of the genome considers the emerging three-dimensional studies that regulatory elements might have the capacity to act in *trans* to affect genes on other chromosomes. As painted by the picture of the above examples, the transcriptional instructions given by enhancers to promoters may be delivered from a different chromosome with the looping out of intervening DNA. Although it is likely that the nucleus is governed by self-organizing principles in regards to the chromatin context and transcriptional status of chromosomes, there is a wide-range of probabilistic intra-chromosomal and inter-chromosomal associations, and accordingly, further genome-wide experiments are necessary in order to achieve a complete paradigm

of gene regulation in mammals. Currently, experimental data is lacking that proves inter-chromosomal gene regulation as another level of control within the nucleus, and as a result, the concept remains highly contentious.

With this in view, one way to identify enhancers in the genome is by their sensitivity to digestion by small amounts of DNase I. Transcriptionally active regions are associated with the presence of DNase I hypersensitive sites in the genome since DNA binding sites at these areas are more accessible to transcription factors and accessory proteins due to their nucleosome depletion (Gross and Garrard, 1988). Hence, DNase I hypersensitivity sites (DHSs) are highly cell specific and indicate regions of open chromatin that regulate gene expression through binding of transcription factors (Sheffield et al., 2013). A high-throughput method exists to identify DNase I hypersensitive sites across the entire genome using deep sequencing (Crawford et al., 2006). Though DNase-seq is not a substitute for ChIP-seq analysis, DNase-seq is valuable in identifying active enhancers and promoters that are independent of prior knowledge of the identity of bound transcription factors (Boyle et al., 2008). It can be effectively used to inventory whole chromosomes for additional *cis*-regulatory elements. Additionally, using this technique, one way to probe for potential *trans*-effects of enhancers is then analyzing DNase hypersensitive cell types in both a wild-type and mutant context. Any potential inter-chromosomal interaction will be detected by the presence or absence of a cluster of hypersensitive peaks in one setting and not the other. Such investigative studies will either add more or less experimental evidence towards the view that enhancers have additional inter-chromosomal gene regulation.

One comment must be made in regard to the inventory of enhancers across the genome and deriving claims for whether a particular enhancer indeed may possess inter-chromosomal gene regulation. All the approaches considered to this point use indirect properties of enhancers and do not examine whether an enhancer directly contributes to gene expression. An enhancer may indeed have inter-chromosomal characteristics, but additional testing, such as the identification of transcription factor binding events or chromatin marks that are representative of function are important to actually understand the biological role of any inter-chromosomal relationship.

Now that the groundwork has been laid and the various levels of regulation have been considered for the central *cis*-regulatory elements in the genome, the specific *cis*-regulatory elements of the *IgH* locus will be explored to understand the purpose of this thesis.

IV. *cis*-Regulatory Elements of the *IgH* Locus

The stage specific development and function of B-cells in the bone marrow that results in an IgM-class immature B-cell is directly dependent on the remodeling events of the immunoglobulin antigen receptor loci, as these rearrangements precisely determine the checkpoints throughout each maturation step (Figure 1). In turn, the cutting and pasting of different antigen receptor genes is a consequence of *cis*-acting regulatory elements and *trans*-acting factors. While a number of *cis*-regulatory elements have been located throughout the ~3 Mb *IgH* locus on mouse chromosome 12, specific attention here will first be given to those significant elements within the 270 kb D_H to C_H region,

followed by a selection of elements elsewhere on the *IgH* locus, as the pro-B cells prepare for *IgH* V(D)J recombination (Figure 5).

Two major *cis*-regulatory elements are currently known to exist within the D_H to C_H region on the *IgH* locus that are marked by lineage-specific DNase I hypersensitive sites. This includes the most 3' D-gene segment, DQ52, which has both promoter and enhancer activity (Kottmann et al., 1994), and the 700-bp Eμ intronic enhancer that is 3' of the last J-gene segment. The overview here will start with those *cis*-regulatory features of DQ52 promoter, known as P_{DQ52}, and its flanking gene partner, DFL16.1, and their role in gene regulation as it relates to chromatin, transcription, and recombination.

Initial chromatin analysis on the *IgH* locus in pro-B cells demonstrated a tissue-specific restricted and patterned environment for histone acetylation to the D_H to C_H region and not to the V_H genes (Morshead et al., 2003). Even the chromatin mark H3K4me3, which is a footprint for active promoters, demonstrated little enrichment in the V region with several peaks of enrichment in intergenic regions (Malin et al., 2010). Therefore, the specific architecture of active and inactive chromatin marks in the D_H to C_H region is unique, insofar as it points to a functional developmental significance for this area in mouse pro-B cells. The character of these active histone marks H3K9ac and H3K4me3 were found to be further restricted to nucleosomes of the outermost 5' and 3' genes, DFL16.1 (chromosome 12:114,720,388) and DQ52 (chromosome 12: 114,668,722), respectively (Chakraborty et al., 2007) (Figure 6). These two genes are separated by a distance of ~52 kb that contains the D-gene family. All D-genes between these boundaries share a high degree of homology within a repeated 4 kb sequence and are known as the DSP gene segments. This intervening region is masked with H3K9me2

heterochromatin marks (Bolland et al., 2007; Chakraborty et al., 2007). While the two flanking genes, DFL16.1 and DQ52, were most sensitive to DNase I digestion, the intervening segments were insensitive. In addition, the active chromatin pattern at DQ52 has a dramatic increase in both H3K9ac and H3K4me3 less than 1 kb 5' from the J_H cluster of genes in pro-B cells (Chakraborty et al., 2009). This hierarchal pattern that has specific boundaries of active chromatin peaks followed by a dramatic elevation in the J_H family is significant since it points to locations where the RAG nuclease likely prefers association to prepare and initiate V(D)J rearrangement events on the *IgH* locus.

Connected to this, upon targeted deletion of the J_H gene segment, direct V_H to D_H rearrangement is not observed (Chen et al., 1993). Therefore, the chromatin picture is instructive for how the complex process of RAG-mediated rearrangements take place. Even within the overall paradigm of V_H to DJ_H rearrangement, a V_H gene segment must overcome the intervening heterochromatic-dominated D_H segments to join a recombined DJ_H segment. This raises questions about the outermost D_H gene, DFL16.1, which displays active chromatin marks and transcription characteristics and appears to function like a castle's most exterior wall that guards the king and queen inside its gates. Technically though, DFL16.1 is not the outermost D_H gene, as this distinction belongs to DST4.2, which is located 43 kb upstream of DFL16.1 and although it has an RSS sequence, it's never been observed to have undergone D to J rearrangement and also no transcripts have been reported from this site (Featherstone et al., 2010; Ye J., 2004). Returning to DFL16.1, this gene has a bidirectional promoter and its transcript level compares to DSP2, and when compared to its flanking partner, the level is about 50% of DQ52 (Chakraborty et al., 2007). Because of DFL16.1's isolated position, several

studies have led to speculations that the V_H – D intergenic region contains additional *cis*-regulatory elements and that DFL16.1 marks a chromatin boundary.

In the first study, a distal V_H gene was relocated several megabases and placed 1 kb 5' of DFL16.1 in pro-B cells (Bates et al., 2007). Interestingly, the relocated V_H gene underwent direct rearrangement to unrearranged D_H gene segments. This led to the conclusion that the chromosomal location of a V_H gene is the determining factor as to whether it's used or not for V(D)J recombination. Also, it was thought that there may be a *cis*-regulatory element in the V_H-D intergenic region that was able to influence V_H to DJ_H recombination (Bates et al., 2007). In a second study, a germline mutation was made in mice that removed the whole 100 kb intergenic region between V_{H81X} and DFL16.1 and placed $V_HQ52.2.4$ adjacent to the D_H locus (Giallourakis et al., 2010). This configuration of the locus led to elevated levels of antisense D transcripts that started within the D locus and went upstream through the D locus and continued until $V_HQ52.2.4$. This antisense transcription was reflected by an increase in D to J_H rearrangement and proximal V_H to DJ_H gene recombination in T cells. This result led the authors to conclude that the deletion of the intergenic region removed a negative element, specifically two insulators, which act to suppress anti-sense D transcription. By removing these insulators and physically placing the V_H gene next to the D genes, it triggered antisense D_H transcription which resulted in activate rearrangement of D_H to V_H genes (Giallourakis et al., 2010). This was supported by the identification of an assemblage of three DNase hypersensitivity sites that were located near DFL16.1 and have been shown to have an insulator sequence that contains CTCF-binding elements (Featherstone et al., 2010) (Figure 5). The purpose of this sequence (CCCTC) is to serve

as a barrier element to recombination and prevent spurious germline rearrangements between V_H, D_H, and J_H gene segments. Collectively, these three studies pointed to additional *cis*-acting elements directly upstream of DFL16.1 that will be discussed further below.

Opposite of DFL16.1 on the 3' end in the D_H domain is DQ52. Non-coding RNA transcription on the germline allele, also called sterile transcription, from the DQ52 promoter, P_{DQ52}, reflects in part the active chromatin status of this gene in pro-B cells. The promoter P_{DQ52} becomes active prior to the initial D to J_H recombination event and initiates the μ0 transcript (Alessandrini, A., and Desiderio, 1991). This transcript is directed away from the J_H and Cμ exons in the antisense orientation and gets spliced to the J_H1 splice donor site (Schlissel et al., 1991; Chakraborty et al., 2007). After D to J rearrangement, the DJ joined segment produces Dμ transcripts, which is followed by non-coding RNA transcription in the V_H locus (Reth and Alt, 1984; Yancopoulos and Alt, 1985; Corcoran et al., 1998). Every D_H gene segment that is located upstream of DQ52 has a bi-directional promoter, where at D to J_H rearrangement, are activated to produce both a sense and antisense transcript towards and away from the Cμ exon. The anti-sense transcript level decreases rapidly in a distance-dependent manner and the sense transcript is spliced so that the rearranged DJ_H segment is joined to the Cμ exon.

The accumulation of evidence and patterns from active chromatin marks and transcription generation from the P_{DQ52} promoter has the functional consequence of preference for specific D to J_H gene rearrangement over others, as DQ52 to J_H recombination takes place at a higher frequency than the intervening DSP gene segments. Though higher than the DSPs, some studies accounted for DFL16.1 being used at even a

higher frequency than DQ52 in mature B cells (Chang et al., 1992; Tsukada et al., 1990). Although these analyses were executed in mature B cells that may have already had secondary rearrangements or other chromatin based mechanisms of preference selection towards the flanking DQ52 and DFL16.1 flanking genes, analysis in pro-B cells has confirmed the preference for these outer segments in rearrangements and also similar usage levels (Tsukada et al., 1990). As a result of these rearrangement preferences, the best mode of interpretation for these choices is that these outer-gene segments, due to their active chromatin marks, especially H3K4me3, are most accessible to recombinase in pro-B cells (Figure 6). This conclusion is corroborated with data on four levels that serve to explain the constraints for ordered V(D)J rearrangement. First, the presence of active histone marks in a distinguishable pattern, and the lack of these marks in V_H genes explains why the initial rearrangement takes place between D_H to J_H genes and excludes V_H genes. The low enrichment of active histone marks on V_H genes in pro-B cells renders them inaccessible to RAG (Chowdhury, D. and Sen, R., 2001).

Second, it is the J_H region that has the highest level of RAG proteins in the *IgH* locus at the pro-B cell stage, as RAG is not detectable in V_H gene segments (Ji et al., 2010; Subrahmanyam et al., 2012) (Figure 6). Consequently, there is not simply a preference for a patterned chromatin landscape in pro-B cells, but this pattern provides focal establishment of recombination centers that give direct accessibility for RAG to carry out the gene rearrangement in a precise manner. To further explain the preference for certain D_H gene segments over others, a special spatial configuration has been observed where the 3' domain of the *IgH* locus undergoes a conformation change and compacts as it folds into a three-loop structure that locates the flanking D_H gene segments

in close proximity to the RAG-bound recombination centers (Guo et al., 2011). Such a looping structure explains why the flanking D_H gene segments are used at a higher frequency than the intervening segments, for the D_H-associated RSSs are brought into close proximity at the initial D to J_H rearrangement step. Taken altogether, the enhancer and promoter activities of the *cis*-regulatory element P_{DQ52} to generate the μ0 transcript, including the specific active chromatin boundaries of the flanking DFL16.1 and DQ52 genes, along with the increase in these marks in the J_H gene segments, and the spatial conformation of these flanking D_H gene segments to the recombination center, helps to explain their higher usage for recombination frequency, the precise localization of RAG to this area on the *IgH* locus, and ability to govern the initial D to J_H rearrangement step.

The functional consequence of this paradigm is that the intervening DSP gene segments recombine less frequently, and this can be explained on two basic levels. First, these gene segments are dominated by heterochromatin marks that make RAG inaccessible, as RAG binding is specific to H3K4me3 (Figure 6). Second, according to the dynamic looping model of the 3' *IgH* locus, the DSP gene segments are further away from the recombination center, and therefore are not used as much in gene rearrangements as the flanking regions. The DSP gene segments are still used in rearrangements, but at much lower frequency and only when the dynamic looping allows a transient movement of the loop such that the RSSs of specific DSP gene segments are brought near the RAG recombination center in the J_H cluster. Yet, in order to precisely explain the impact of *cis*-regulation on the *IgH* locus, the critical significance of the Eμ enhancer and its role in gene regulation must be outlined. Deletion studies from DQ52

provided an important clue for the significance of the Eμ enhancer and later studies affirmed its vital role in *IgH* rearrangements in pro-B cells.

One important investigation implemented a targeted mutation in the P_{DQ52} promoter. No significant impact on D to J_H rearrangement was observed in pro-B cells and μ0 transcripts were still detected (Afshar et al., 2006, Nitschke et al., 2001). This suggested that the heterologous P_{DQ52} promoter was not completely diminished in this context. Yet, these studies were significant because it pointed to an element outside of DQ52 that is responsible for initial D to J_H rearrangement gene rearrangement. When investigators replaced the 700-bp Eμ intronic enhancer *cis*-regulatory element (chr12:114,665,501 - 114,666,175) with a phosphoglycerate kinase promoter – neomycin resistant gene cassette (PGK-Neo^R), it resulted in complete inhibition of both μ0 transcripts and D to J_H rearrangement (Perlot et al., 2005). Even removal of the 220 bp core region of the Eμ enhancer resulted in a significant reduction in D to J_H (Perlot et al., 2005; Chakraborty et al., 2009). When both the P_{DQ52} promoter and the full 700 bp Eμ enhancer are deleted, the results are similar to the deletion of the 220 bp core region of the Eμ enhancer (Afshar et al., 2006). Therefore, DQ52 is dependent on the Eμ enhancer and this was substantiated by a critical DNase hypersensitivity experiment, which showed that when the core of the Eμ enhancer is deleted, the hypersensitivity peak at DQ52 is also removed (Chakraborty et al., 2009). Collectively, these results have pointed to the critical role for the *cis*-regulatory Eμ intronic enhancer in activating the *IgH* locus for gene recombination in B-cell development. The significance of this enhancer is further demonstrated from chromatin and transcription analysis (Chakraborty et al., 2009; Chakraborty et al., 2007).

At the level of chromatin, the effect of the Eμ enhancer to the D_H to C_H region has a unique pattern that is distinguished from the pattern noted previously for DQ52 DFL16.1 genes (Figure 6). This effect was explored by deleting the 220 bp "Eμ core" in pro-B cells (Chakraborty et al., 2009). While the flanking D-gene segments have the active chromatin marks that are then sandwiched by a distance of 52 kilobases (kb) of intervening heterochromatin, there is a specific cluster of active acetylation marks present in the J_H genes that is significantly diminished in the absence of the Eμ core (Chakraborty et al., 2009). This reduction in active marks is complemented with an increase in inactive marks; specifically, there is a significant increase in H3K9me2 at the J_H gene cluster in the Eμ-core deleted setting. There is also a reduction in H3K9ac and H3K4me3 throughout the D_H to C_H region. This observation includes a major lowering of active chromatin marks at both DFL16.1 and DQ52 in the deleted Eμ core setting and as far as 15 kb downstream of the Eμ enhancer. Nonetheless, not all areas in this region of the *IgH* locus appeared Eμ dependent. In the deleted Eμ core setting, there was less reduction observed between DQ52 and J_H4 for H3K4me2, and the gene region between DQ52 and Cμ had a comparatively reduced level H3K9me2 regardless of whether Eμ is present or absent. Such data indicated there are Eμ dependent and Eμ independent events as the *IgH* gene segments prepare for rearrangement (Chakraborty et al., 2009). Regulation of these independent events are likely attributed to other *cis*-acting elements on the *IgH* locus that are further explored below. In view of the Eμ dependent chromatin changes between the D_H to C_H region of the *IgH* locus, the key conclusions from this pioneering chromatin study were twofold. First, deletion of the Eμ core acted locally to affect the chromatin structure of genes in the J_H cluster that are between 650 bp and 2 kb

away, and as a result, the heightened elevation of histone acetylation and H3K4me3 over the J_H cluster is Eμ dependent. Second, deletion of the Eμ core led to the reduction of chromatin at genes 3 kb (DQ52) and 54 kb (DFL16.1) upstream and 15 kb downstream. As a result, insofar as the histone marks for H3K9me2, H3K9ac, and H3K4me3, the Eμ core had an effect that was both local, extending a total distance of 69 kb with respect to the chromatin landscape. The importance of this chromatin view is that euchromatin is limited to distinct regions that are controlled by the Eμ enhancer, and this paradigm then limits the regions of RAG binding, thereby restricting the preparation and initiation of D_H to J_H gene rearrangement in pro-B cells. In order to initiate the dynamic chromatin changes in pro-B cells for gene rearrangements, may be the role of putative "pioneering factors," perhaps noncoding RNAs that are known to modulate histone tails, which initiate intergenic transcription within silent chromatin and such transcription then promotes histone exchange for active modifications (Mito et al., 2005; Orphanides and Reinberg, 2000). Transcription at the Eμ enhancer is then critical to the ordered process of gene rearrangement events in pro-B cells as it reflects, in part, the chromatin features.

Although the Eμ enhancer functions as the promoter for Iμ sterile transcription of non-coding RNA in germline alleles (also termed iEμ), and this has been restricted to the 700 bp sequence between XbaI and EcoRI in the J_H and Cμ area of the *IgH* locus, it is the 220 bp Eμ core where the majority of binding sites for transcription factors exist (Nikolajczyk et al., 1999; Su and Kadesch, 1990). This Eμ core is surrounded on each side with matrix attachment regions (MARs) that have yet to demonstrate any role in transcription or recombination (Jenuwein et al., 1997; Sakai et al., 1999). At the level of transcription, deletion of the 700 bp Eμ, including its core and MARs, produced results

that demonstrate its primary role as a *cis*-regulatory element. While antisense intergenic transcription occurs throughout the D_H and J_H gene clusters prior to recombination, deletion of Eμ resulted in loss of iEμ and D antisense transcription over a 60 kb region of D_H segments (Afshar et al., 2006; Bolland et al., 2007). This result supported the role of Eμ as the regulator for D-J rearrangements by its ability to activate antisense transcription through the D domain. The Eμ enhancer has also been recognized to have ongoing transcription throughout V(D)J rearrangements in both pro-B cells and B-cell development in the bone marrow (Bolland et al., 2004; Osborne et al., 2007). In this regard, the non-coding Eμ enhancer has been suggested to be the first "super-intergene" ever found (Fraser, 2006; Matheson and Corcoran, 2012). For, it has all the dressings of a significant *cis*-regulatory element with its featured MARs, core transcription factor binding region, and precise location adjacent to the indispensable J_H gene family.

All of these transcription roles were further refined in the deleted Eμ core context (Chakraborty et al., 2009), where germline transcription was observed to affect V_H genes up to 450 kb upstream. However, as mentioned earlier, these same V_H genes bear no marks for either H3K4me3 or H3K9ac, suggesting that transcription at these gene promoters is not under the control of epigenetic regulation. Also in the deleted Eμ core setting, all transcripts in the Cμ area were reduced by approximately 7-10 fold and D_H amplicons were also reduced, but in a much higher amount (~10-50 fold). Alongside this, RNA polymerase II binding was also significantly reduced. Hence, the transcription studies on the deleted Eμ core demonstrate that one function of Eμ is to attract RNA polymerase II in the presence of the open chromatin context. Second, the Eμ enhancer

has specific and selective transcription targets in the D_H to C_H region that is related to the chromatin structure, but is not epigenetically related in the V_H region.

The pattern between open chromatin and transcription in the D_H to C_H region is also supported by DNase I digestion experiments with one caveat. In Eμ sufficient pro-B cells, the flanking D_H genes DFL16.1 and DQ52 demonstrate DNase I sensitivity while the intervening region is insensitive. Rapid degradation of amplicons is also observed in the 7.8 kb region that is between DQ52 and Cμ, which is indicative of DNase I sensitivity. These patterns of sensitivity are reflective of the unique chromatin structure, and parallels the heightened presence of H3K4me3 marks in comparison to the rest of the locus. Yet, in the deleted Eμ core model, there is minimal degradation at DFL16.1 (Chakraborty et al., 2009). The most likely explanation of this phenomenon is that DFL16.1 is regulated by other *cis*-regulatory elements in addition to the Eμ enhancer as the *IgH* locus prepares for gene rearrangement. Such regions may include the *IgH* 3'RR, a 40 kb long regulatory region located downstream of Cα, and potential *cis*-regulatory elements within the 5' V region of the *IgH* locus (Figure 5).

The *IgH* 3'RR was initially identified by the presence of a DNA sequence that had B-cell specific enhancer characteristics located ~25 kb downstream of the most 3' C_H genes, Cα (Pettersson et al., 1990). Several additional enhancers were identified by the presence of DNase I hypersensitive sites and the activity of these enhancers were typically confirmed in transient transfection assays using B cell lines that reflected different stages in B-cell development (Michaelson et al., 1995). In total, four enhancers (hs3a, hs1.2, hsb, and hs4) and four additional DNase I hypersensitive sites were identified, known as hs5-8 (Garrett et al., 2005) (Figure 5). One special feature of hs5-8

is that it contains a CTCF/Pax5 binding region (Chatterjee et al., 2011). As a result, the *IgH* 3'RR contains two related, but distinct segments. The first is the ~28 kb region consisting of the four enhancers, and the second is the ~12 kb region. Experimental analysis of targeted mutations in mice from several labs demonstrated that the *IgH* 3'RR enhancers are critical for somatic hypermutation and class switch recombination, but not for V(D)J joining (Rouaud et al., 2012). Likewise, mild phenotypic changes were also observed when an 8 kb region of the hs5-7 CTCF/Pax5 sites were deleted (Volpi et al., 2012). The lack of any major changes in deleting portions of the 3'RR might be attributed to the presence of hs8, which contains two CTCF sites and was not included in the hs5-7 deletion studies. Therefore, future deletion studies that include hs8 may indeed reveal a greater regulatory role for V(D)J recombination.

While deletion studies have not revealed a significant role for the 3'RR with respect to V(D)J recombination, 3C studies in MPC11 plasma cells, variant and normal splenic B cells, and resting B cells have shown that the 3'RR does interact with the J_H - Eμ region in both Eμ sufficient and Eμ deficient cells (Ju et al., 2007; Wuerffel et al., 2007). This interaction may explain the Eμ independent events that are observed with respect to H3K4me2 and H3K9me2 in the DQ52 to Cμ region of the *IgH* locus. In this model, the 3'RR has two roles. First it leads to demethylation of H3K9me2, and second, the 3'RR delivers the histone acetyltransferases to the H3K4me2 marked promoter for transcription initiation, which results in H3K4me3 histone modifications and an accessible state for recombinase. The meaning of this interaction between the 3'RR and *IgH* locus is likely predicated on the location of CTCF sites on the *IgH* locus. Such sites have been located in two places: (1) 3.1 and 5.7 kilobases 5' of DFL16.1, known as the

IGCR1, and (2) several locations in the V_H region (Ebert et al., 2011; Featherstone et al., 2010; Guo et al., 2011). Of particular importance, it has been demonstrated that the insulator 3'RR CTCF/Pax5 hs5-8 binding region interacts with the DFL16.1 proximal 5' CTCF sites (IGCR1) (Guo et al., 2011). Although deletion of the IGCR1 demonstrated a key role for the IGCR1 in terms of the balance between proximal and distal V_H rearrangements, it is not required for overall V_H to DJ_H recombination (Guo et al., 2011). A future study that deletes both the IGCR1 and the 3'RR may yield a larger defect in V_H to DJ_H rearrangements, however the initial D to J_H recombination event is unlikely to be affected.

Consequently, the exact role of these *cis*-regulatory elements, 3'RR and the IGCR1, and how they relate to DFL16.1 and other events in pro-B cells, remains unclear. These regions are likely interacting to prepare the *IgH* locus in pro-B cells for gene recombination and assist in the control of other events in B-cell development and maturation. Chromatin looping studies have demonstrated the *IgH* domain undergoes contraction and decontraction events to influence partner selection in developing B cells for gene recombination. In pro-B cells, the *IgH* locus contracts to bring the separate gene clusters close together for V_H to DJ_H recombination. This contraction is followed by a locus decontraction at the next stage of development to prevent rearrangements on the second allele. Several labs have used chromosome capture studies (3C and 4C) to view several large-scale contraction phenomena in pro-B cells (Guo et al., 2011; Medvedovic et al., 2013; Sexton et al., 2009; Zhao et al., 2006). These investigations have demonstrated several important features regarding flexible long-range interactions between *cis*-regulatory components on the *IgH* locus. First, interactions have been

demonstrated between the 3'RR and V_H genes in Rag-/- pro-B cells prior to V(D)J joining (Medvedovic et al., 2013). These communications we're found to depend on Pax5, YY1, and CTCF. Interestingly, such an interaction might indeed be a preparatory feature of the *IgH* locus in pro-B cells as it is maintained when *cis*-regulatory features, such as Eμ, IGCR1, and portions of the 3'RR are deleted.

Several other interactions between *cis*-acting elements that correspond to CTCF bound V_H sites have been found in the 5' V_H region (Guo et al., 2011). These loops are reported to extend between 30-500 kb between different V_H regions, and even one set of loops associates with DFL16.1 in an Eμ independent manner. This may explain how the mild hypersensitivity peak at DFL16.1 is maintained in Eμ deficient pro-B cells and also the preference for DFL16.1 rearrangements with proximal V_H gene segments (Perlot et al., 2005). Taken altogether, the *IgH* locus then positions itself early in pro-B cells for the main recombination event, and there is an organization within the locus that is broadly maintained independent of the *cis*-acting elements, including Eμ.

Second, when J_HT-/- Rag2-/- Eμ-/- pro-B cells are compared to Rag2-/- pro-B cells with respect to locus contraction, a significant role for the Eμ enhancer is observed at the 3' proximal end of the locus. In this setting, the dynamic folding of the locus is dependent on Eμ as it interacts with the V_H region, several kb 5' of DFL16.1, and also with the 3'RR (Guo et al., 2011). The Eμ enhancer was described as the critical *cis*-regulatory element that forced the three-dimensional structure into its unique fold for V(D)J recombination. Therefore, the Eμ *cis*-regulatory element is then able to position segments of the *IgH* locus to a nuclear environment that is accessible to transcription factories, including RAG recombinase.

Given the need for stable proximity of the locus during gene rearrangements, ncRNAs have been suggested as a mechanism that modulates the folding of V, D, and J gene segments so that the V chromatin domains interact with the DJ cluster at precisely the correct time (Matheson and Corcoran, 2012; Jhunjhunwala et al., 2009). Noncoding RNAs are known to nucleate structure. MARs are known to bind proteins associated with repression and activation and have also been established as organizers that maintain chromatin into loops to keep the associated DNA with the nuclear matrix. MARs flank the core of the Eμ enhancer and are also deposited at many places throughout the V_H region. In this model, Eμ through its MARS, recruits the D and J gene segments into a nuclear space that is favorable for D to J_H recombination, and also where non-coding RNA transcription takes place to allow the non-coding V region to associate with the transcription factory to enable V to DJ_H joining. In summary, herein lies the pivotal role for Eμ enhancer: control the D_H to J_H chromatin landscape to initiate site specific *IgH* gene rearrangement for optimal production of a function B-cell receptor.

Other *cis*-regulatory features may be present in the V_H domain of the *IgH* locus (Figure 5). This comes from the recognition that the 5' V_H cluster is not deleted during the process of V(D)J recombination. Interactions through radial repositioning between the V_H domain and other regions of the *IgH* locus are likely necessary for accessibility control, allelic exclusion, and the temporal positioning of RAG nuclease during B-cell development and maturation. One of the initial lines of evidence that supported this view was the identification of several DNase I hypersensitive peaks centered around Zfp386, a Kruppel-like zinc finger protein, located 30 kb upstream of the last 5' V_H gene

(Pawlitzky et al., 2006). Further analysis on this site featured no relationship to allelic exclusion or gene recombination.

Another line of evidence for *cis*-acting elements in the V_H gene segment comes from the presence of Pax5 binding sites in this region (Ebert et al., 2011; Zhang et al., 2006). Although Pax5-/- mice have no major defects in any of the ordered recombination steps in developing B cells (Hesslein et al., 2003; Nutt et al., 1997), Pax5 may direct RAG recombinase to many areas in the V_H domain, including the V_H3609 gene family (Ebert et al., 2011; Subrahmanyam and Sen, 2012). These Pax5 binding sites also bind a collection of proteins all implicated in DNA looping in pro-B cells: CTCF, E2A, YY1 and Rad21, an integral subunit of the cohesin complex. These conserved binding sites have been called "PAIR" elements (Pax5-associated intergenic repeat) (Ebert et al., 2013). PAIR elements have been identified in 14 locations, and 11 of these are upstream of V_H3609 in the distal V_H domain. These *cis*-acting elements have been shown to operate as promoters and give rise to anti-sense transcription in pro-B cells. Due to their location in the distal V_H region, they have been implicated in Pax5 dependent *IgH* contraction and V_H to DJ_H rearrangements.

In conclusion, this sub-section has highlighted the significant role of *cis*-acting elements as they have been described in the genome and also those that are specific to the *IgH* locus. Particular focus was given to the β-globin locus LCR, numerous interactions reported between *cis*-elements and other chromosomes, and the human insulin locus was highlighted to demonstrate *cis*-regulation of a gene at a considerable distance due to the chromatin pattern. However, many questions for these *cis*-regulatory components of the genome remain, as several studies have left open roles for other *cis*-acting factors that

further explain the *in vivo* interaction. Next, the significant *cis*-regulatory features of the *IgH* locus were described. While many elements have been identified, and all likely have a role in B-cell development that need further exploration, the Eμ enhancer was demonstrated to have distinct features in regulating chromatin, transcription, and recombination. Although the Eμ enhancer has a definite critical role in D to J_H gene rearrangements in pro-B cells, it is important to study whether the Eμ enhancer has gene regulatory features that extend beyond its proximal domain and into other areas of the genome. Moreover, the specific characteristics of the Eμ enhancer lead to the consideration of whether or not this enhancer retains its *cis*-regulatory function as it relates to the D-J_H domain of the *IgH* locus when it is moved to other areas in the *IgH* locus. The central prediction from *in vitro* transfections assays is that the enhancer will retain its ability to regulate gene expression regardless of distance from their promoters. Both projects in this thesis will study these hypotheses and test whether the Eμ enhancer affects other genes within and beyond its proximal domain, and if the Eμ enhancer retains its *cis*-regulatory characteristics, with respect to chromatin, transcription and recombination, when it is relocated to other regions in the *IgH* locus. In this regard, two brief sub-sections follow that describe the background for key experiments used in this thesis to test both the relation of the Eμ enhancer's location to gene regulation, and whether the Eμ enhancer affects other genes beyond the D to J_H region.

V. Use of Bacterial Artificial Chromosomes to study *cis*-regulatory features of the *IgH* locus

One major goal in developmental biology is to study tissue-specific regulatory elements within their native context. Prior to the Human Genome Project, most

functional studies to investigate gene regulatory elements were accomplished by sub-cloning small fragments of genomic DNA into various reporter vectors. This included transfection assays, where different combinations of *cis*-regulatory elements were inserted into plasmids and used for transfection assays to study the transcriptional effects of a given set of regulatory elements on a gene of interest. Using such methods, enhancers were originally defined as a region of DNA that is able to activate transcription from a target promoter in both a location and orientation independent manner (Banerji et al., 1983; Gillies et al., 1983). In fact the Eμ intronic enhancer was initially discovered as a transcriptional enhancer in this way, as it was observed to drive the expression of immunoglobulin or other transgenes in an orientation and position independent manner in lymphoid cells alone (Banerji et al. 1983; Gilles et al., 1983; Neuberger 1983). Yet, it has been increasingly clear that studying a tissue-specific enhancer outside of its natural context in transgenic reporter assays may not reflect the intrinsic character of that *cis*-regulatory element in its own natural genetic environment (Chatterjee et al., 2012).

Bacterial artificial chromosomes (BACs) overcome this defect as they may be used to study regulatory elements located distantly from a gene. BACs are single copy replicons originated from the naturally occurring *Escherichia Coli* fertility plasmid (F-) and were first developed as large insert clones to assist in the construction of a consistent set of overlapping clones for the Human Genome Project (Shizuya et al., 1992). Due to their large insert capacity, ranging from 100-750 kb of genomic sequence, BACs are able to be comprised of whole genes and the necessary *cis*-regulatory elements, and at the same time be flanked by regulatory DNA elements that provide the signals needed for correct spatio-temporal gene expression (Antoch et al., 1997; Sharan et al., 2009; Valjent

et al., 2009). In addition, BAC transgenes are able to mimic the natural expression of the gene(s) of interest to a much higher degree than conventional transgenes (Gong et al., 2003; Yang et al., 1997). The original basis for using BACs to study genome structure and function came from a mouse study where a ~100 kb construct was introduced into the germline of a mouse and demonstrated the expected tissue-specific expression pattern (Kim et al., 1996). Along with this, BACs are able to be used to study the developmental consequence of accurate but excess expression of genes. One critical feature of BACs is their ease of manipulation, and therefore, not only can the expression of genes be studied, but the deletion, modification, or insertion of different *cis*-regulatory elements allows BACs to be used as means to study mechanistic and functional questions. Therefore, BACs represent an ideal tool to study the manipulation of *cis*-regulatory sequences within the native genome environment, and thereby allows further investigation about *cis*-regulatory features that cannot be detected in *in-vitro* transfection assays. With respect to the Eμ enhancer, BAC sequences are able to be manipulated so that, while keeping the Ig D_H and J_H locus intact, the location of the Eμ enhancer may be different from the endogenous position. This system sets up the experimental investigation as to whether the Eμ enhancer is able to regulate genes at different locations as it does in the endogenous setting.

VI. *Cis*-regulatory elements and probing for other interactions

One of the main mechanisms that determine tissue specific patterns of genome organization is the regulation of differential gene expression. In complex genomes, *cis*-regulatory elements are distributed in a variety of intergenic and intragenic locations that

are hundreds of kilobases away from their target genes (Nobrega et al., 2003). The purpose of these *cis*-acting elements is to fine tune the expression of genes in a cell-type specific manner. Enhancers are one of the significant *cis*-acting elements that control tissue specific gene expression. Essentially, they are the main drivers of gene expression. Enhancers recruit transcription factors that in turn disrupt histone continuity and establish nucleosome-depleted regions that further attract ATP-dependent chromatin remodeling complexes to open the chromatin structure for enhancer coordinated gene function. Genome-wide mapping data on nucleosome occupancy indicates that at *cis*-regulatory elements, such histone replacement closely reflects the presence of nuclease-hypersensitive sites (Mito et al., 2007). As the key that unlocks the door of gene regulation, the ability for any enhancer to affect other *cis*-regulatory elements in intergenic or intragenic regions may be probed by their sensitivity to minimal amounts of DNase I digestion.

X-ray structural analyses of DNase I complexes showed the protein binds only to the minor groove and the backbone phosphate group to cleave the sequence in neither a sequence nor base-specific manner (Suck, 1994). There is no contact between DNase I and the major groove of right handed DNA duplexes. Such binding of DNase I then alters the conformation of the groove geometry and distorts the B-DNA conformation (Lahm and Suck, 1991; Weston et al. 1992). The cutting rate of DNase I is sequence dependent and the cleavage probabilities are low where a string of As and Ts are present and also in GC rich areas (Suck, 1994).

Originally isolated from bovine pancreas by Kunitz in 1940, the DNase I gene is located on mouse chromosome 16 at position 4,037,101-4,040,024. Spanning nearly 3 kb

of DNA, the gene consists of eight introns and eight exons (Peitsch et al., 1995). It is ubiquitously expressed in mammalian tissues and has many roles, including the removal of DNA from nuclear antigens (Oliveri et al., 2004). Mice that are deficient in DNase I demonstrate autoimmune characteristics. It also operates as a transcription factor that is implicated in Fas expression on the cell surface in human cells (Oliveri et al., 2004). When used as a structural probe to map regulatory DNA, the DNase I digested fragment length parallels transcription factor occupancy (Lazarovici et al., 2013). Nucleosome depleted DNA that has bound regulatory factors is more accessible to the DNase endonuclease compared to nearby condensed chromatin that is wrapped twice around a core of four histone pairs. It is this accessibility that marks the hypersensitive sites of a DNA sequence for DNase I digestion. Cleavage takes place at nucleotides that are free of proteins. It is precisely by this means that *cis*-regulatory elements, such as enhancers, can be interrogated and epigenetically marked. One distinguishing feature of chromatin accessibility is the presence multiple DNase-I cleavage sites within a short distance when the enzyme is used on intact nuclei. Therefore, when minimal genome cutting takes place in the presence of the non-specific endonuclease (1%), a systematic picture of the *cis*-regulatory features of the genome emerges. In this framework, regions of the genome that have enhancer-like properties may be inventoried, and even other areas an enhancer may affect within or outside a chromosome can be investigated. As a result, this tool can be used with the Eμ enhancer to probe for other sites of potential regulation, which will be further described below.

Transition Remarks

From the above analysis, several conclusions may be drawn concerning the *cis*-regulatory elements on the *IgH* locus that set up the projects for this thesis. First, the discovery of the Eμ enhancer in transfection assays did not consider the complex cellular environment of multiple *cis*-regulatory elements that interact to control gene regulation (Banerji et al., 1983; Gillies et al., 1983). For over thirty years, the study of enhancers that takes into consideration the *in vivo* chromatin environment, other *cis*-elements, and an enhancer's location has been difficult due to the lack of technology to probe this relationship. Although enhancers were classically defined as *cis*-regulatory elements that are able to activate transcription independent of location and distance, it has not been tested in an *in vivo* setting whether this definition remains suitable. BACs provide a tool to express long stretches of gene domains in mice and examine whether enhancers retain their *cis*-acting characteristics independent of distance and location. Specifically, BACs allow an enhancer to be moved into a region dominated by heterochromatin to study whether an enhancer is able to open the nearby chromatin structure and active gene transcription. In addition, BACs allow an enhancer to be placed a considerable distance from their endogenous site to determine whether their gene regulatory features are the same or different. The prediction from *in vitro* plasmid transfection assays is that an enhancer will be able to have the same gene regulatory features independent of distance from the endogenous site in an *in vivo* setting. Chromatin domains were not in view in the original transfection assays, but BACs allow the definition of an enhancer to be further modified to more closely represent an enhancer's endogenous context.

Alongside this, the tissue specific nature of enhancers may be studied using BAC transgenes. As the preceding background stated, the Eμ intronic enhancer

developmentally regulates genes at the pro-B cell stage in B-cell development. Therefore, the study of the Eμ enhancer in a condition specific manner is essential to understanding how this *cis*-acting element regulates genes that lead to the critical function of gene rearrangement, which allows the maturing B-cell to develop and produce a diverse antibody response. Specifically, BAC transgenes allow the Eμ enhancer may to be placed in different locations to precisely determine its ability as a *cis*-acting element to regulate gene rearrangement independent of distance. While other *cis*-regulatory elements have been intensely studied, such as the β-globin LCR, and new ones remain further examination, like the human insulin locus, the tissue-specific effects of the Eμ enhancer are known, including its key position adjacent to the J_H recombination center. Consequently, modified BAC transgenes that include the D_H and J_H domains, and the Eμ enhancer, will allow the testing of the hypothesis that the Eμ enhancer is able to induce gene rearrangements independent of distance and location to the J_H recombination center. In this thesis project, two BAC transgenes were specifically constructed to evaluate the *cis*-regulatory characteristics of the Eμ enhancer's location and distance. In one construct, the Eμ enhancer is relocated into the center of the D_H domain that has a heterochromatin spread of ~50 kb. In this setting, it is of particular interest to test whether the enhancer is able to affect chromatin, transcription, and recombination in the same way as the endogenous site. In addition, when moving the Eμ enhancer into this heterochromatic area, the classical definition of enhancers suggests RAG binding to the recombination center should remain the same. When the Eμ enhancer is moved, if the pattern of RAG binding also changes, then the Eμ enhancer has a direct role in controlling the recombination center in the endogenous setting. This is a significant

experiment, as the ability of a *cis*-acting element to control RAG binding helps to explain RAG recruitment to the J_H domain.

As for the question of an enhancer's distance to control gene regulation, one BAC construct has been designed to relocate the Eμ enhancer at a considerable distance from the endogenous site. Both the above background and other experiments have demonstrated that tissue specific enhancers have a certain distance for *cis*-regulatory function. Previous studies that dissected the Eμ enhancer demonstrated a certain distance where the enhancer directly impacts the chromatin and transcription landscape. From cell line experiments, Eμ impacts chromatin and transcription 54 kb upstream and only 15 kb downstream (Chakraborty et al, 2009). It turns out that the Eμ enhancer is not the only antigen receptor enhancer that impacts gene regulation. Another antigen receptor locus, the β locus of the T-cell receptor is controlled by the Eβ enhancer that is located 3' of Cβ2, and likewise demonstrates a specific distance effect similar to the Eμ enhancer on the *IgH* locus (Oestreich et al., 2006). When histone acetylation is analyzed in Eβ null thymocytes, there is a decrease at the Jβ and Cβ segments that are 20kb away. Similar to the Eμ enhancer, Dβ1-Jβ1 rearrangement requires the cooperation of Eβ which must stimulate rearrangement over a distance of 15 kb. Deletion of Eβ results in a loss of germline transcription that normally precedes any rearrangement, and homozygous Eβ-deleted animals lack αβ T cells (Bories et al., 1996; Bouvier et al., 1996). Likewise, two other antigen receptor enhancers, Eδ and Eα, have also been demonstrated to be long-range developmental regulators of H3 acetylation (McMurry and Krangel, 2000). Collectively, all of these investigations on antigen receptor enhancers demonstrate that these *cis*-regulatory elements are able to positively control active histone marks at

distances related to their functional role in rearrangement. This is a contrast to mouse β-globin LCR studies that have yet to show a relationship between *cis*-acting elements and chromatin structure. In this thesis project, the Eμ enhancer will be relocated a distance of 54 kb upstream from its endogenous site to study its ability to control gene regulation. It is of particular interest to study the issue of distance using a modified BAC construct, and this will directly test the classical definition of enhancers and their ability to activate transcription independent of distance. Coupled to this, studying gene rearrangements in this distant-modified-BAC will lead to further understanding of how the recombination center operates based off the Eμ enhancer's location from the endogenous site.

Collectively, modified BAC constructs provide a tool to study the tissue specific Eμ enhancer in a developmental manner using pro-B cells where the main rearrangement event between the D_H domain and J_H occurs. This system provides a mechanism to test the classical definition of the Eμ enhancer in an *in vivo* setting and either support it or modify it with respect to the chromatin environment. It is of primary importance to further characterize the Eμ enhancer in a condition specific manner because of its critical role in regulating genes in pro-B cells. At the center is the Eμ *cis*-regulatory element, and the prediction of this project is that moving the Eμ enhancer will result in a different chromatin, transcription, and gene rearrangement pattern than the endogenous locus.

Related to the above project is the search for unique *cis*-acting elements both within and outside the *IgH* locus. The preceding background reviewed all the major *cis*-elements reported, including DQ52, Eμ, and the 3'RR. Several DNase I hypersensitivity peaks have been identified throughout the 100 kb intergenic region between the most 3' proximal V_H gene segment and DFL16, and also upstream of the last 5' distal V_H gene

(Figure 5). Nonetheless, besides the PAIR elements, no significant *cis*-acting element has been located in the V_H domain, and the hunt for other *cis*-elements in the 100 kb intergenic region is needed to explain the precise timing and selection of V_H gene rearrangements to DJ_H segments. In addition, beyond the insulators within the IGCR, the identification of additional elements either upstream or downstream of the Eμ enhancer will help explain the existence of other *cis*-acting factors that may have a direct role in pro-B cell development in the bone marrow. As outlined above, DNase I hypersensitivity provides a unique tool to hunt for *cis*-regulatory elements because nucleosome depleted DNA that has bound regulatory factors is more accessible to DNase I compared to nucleosome bound, inaccessible regions. This experiment was modified for B-cells, and was subsequently executed in cell lines to test for Eμ's ability to affect other *cis*-elements both within and outside the *IgH* locus. The experiment was later successfully used in primary B cells from the spleen to analyze changes in DNase I digestion patterns at different time points and samples. Several hurdles existed in the development of this experiment. The first was the inability to produce consistent results using other lab protocols. Protocols were later communicated to be cell-type specific. Coupled to this, some protocols used magnesium (Mg^{+2}) in the digestion buffer. Magnesium speeds up the activity of DNase I in solution, and therefore many experiments were initially slightly over-digested. To overcome this problem, Calcium (Ca^{+2}) was replaced in the digestion buffer to slow down the activity of DNase I to obtain results where only about 1% of the genome was digested.

To test for the role of Eμ to affect other regulatory elements either within or outside the *IgH* locus, three related, but different sets of pro-B cell lines were used. All

three have RAG mutated to maintain the cells in the pro-B cell stage and prevent cell maturation. In the first set of pro-B cell lines, termed "B6D345," the *IgH* allele is in the germline configuration on the C57BL/6 background and lacks endogenous RAG-1, though has transgenic expression of the catalytically inactive RAG-1(D708A) (Ji et al., 2010). The purpose of using this cell line in DNase I studies is twofold. First, it provides a wild-type control for the experiment, and second, since it's on the C57BL\6 background, it then aligns directly to the UCSC Genome Browser after sequencing. The next pro-B cell line used is termed "Rag8-/-." These cells have the *IgH* locus in the germline configuration and are maintained in tissue culture conditions as pro-B cells as they are Rag2-/- deficient. Also, they are different from the B6D345 cell line because they are on the mouse 129 strain background (Simpson et al., 1997). In an identical fashion to these cells, but with one significant alteration, is the third pro-B cell line, termed "Rag2-/- Eμ-/-." This third set has the core domain of the Eμ enhancer deleted, along with the Rag2-/- deficiency (Perlot et al., 2005). As a result, any changes due to the enhancer deletion in the Rag2-/- Eμ-/- cells are observed when comparing to the Rag8-/- cells. While both of these cell lines are on the 129 mouse strain background, differences at the sequencing level can be mitigated with the B6D345 which aligns exactly with the C57BL\6 strain used to generate the alignment in the UCSC genome browser. At the outset, the B6D345 was included due to the unknown nature of the 129 mouse strain aligning with the C57BL\6 strain. In retrospect, no significant differences were observed between these two strain alignments.

In conclusion, this introductory background has effectively set up the central paradigm of this thesis: to study the *cis*-regulatory features of the Eμ enhancer as it

relates to gene regulation (project 1), and to determine whether the Eμ enhancer affects other *cis*-acting elements within the *IgH* locus or outside of it (project 2). The first project will test whether an enhancer's position is necessary for its function, and the second project will interrogate the entire genome for DNase Hypersensitivity peaks that either relate directly or indirectly to the Eμ enhancer. These projects will provide an essential understanding for how the Eμ enhancer gives rise to the genomic events that progress B cells in their development and antibody response.

Figure 1

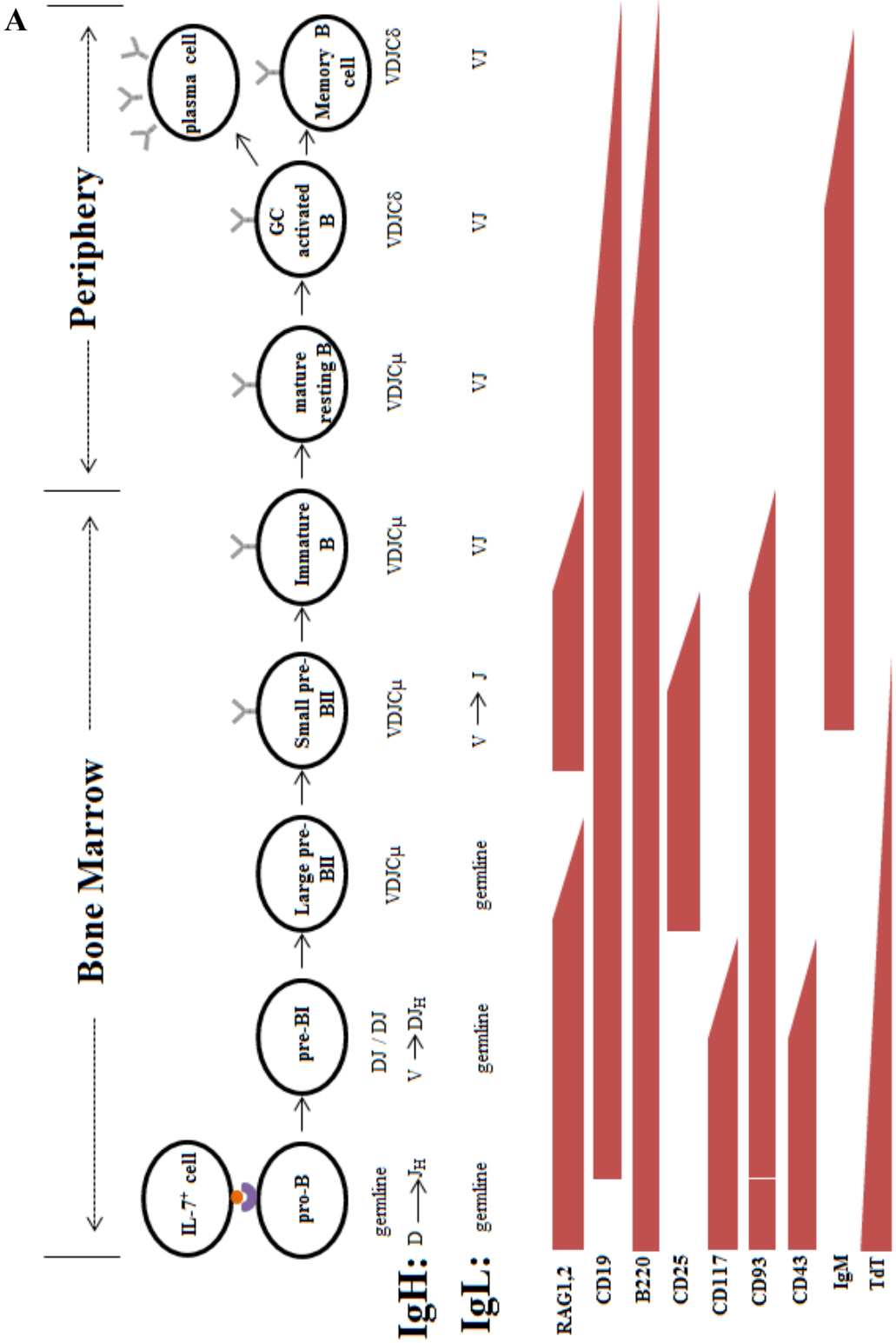
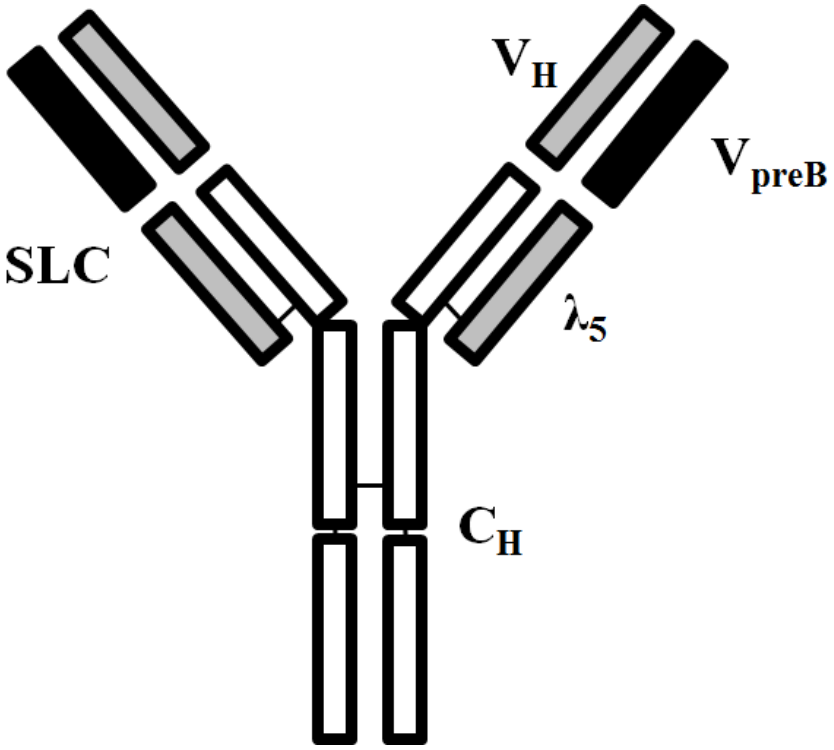

A
Bone Marrow
Periphery
IL-7⁺ cell
pro-B
pre-BI
Large pre-BII
Small pre-BII
Immature B
mature resting B
GC activated B
plasma cell
Memory B cell
IgH: D → JH
IgL:
germline
DJ / DJ
V → DJH
VDJCμ
V → J
VJ
VDJCδ
RAG1,2
CD19
B220
CD25
CD117
CD93
CD43
IgM
TdT

Figure 1

B


V_H
V_{preB}
SLC
λ_5
C_H

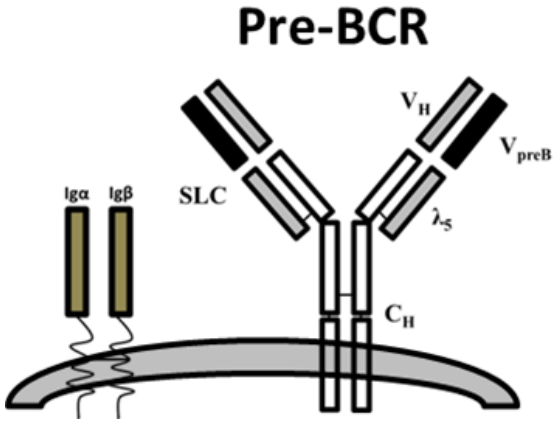

Pre-BCR
V_H
V_{preB}
Igα
Igβ
SLC
λ_5
C_H

Figure 1. B Cell Development, Immunoglobulin Gene Rearrangement and Structure of the Pre-BCR. (A) Schematic illustration of B-Cell development as it relates to immunoglobulin gene rearrangement. Also shown are the CD antigens and other proteins that are expressed at different stages to yield a functional B-cell receptor. After successful heavy and light chain rearrangement in the bone marrow, immature B cells migrate to peripheral organs (spleen, lymph nodes) where maturation is completed. B-cells proliferate extensively in germinal centers where they may undergo class switch recombination and somatic hypermutation to yield antibodies with different effector functions and higher affinity. Some activated B-cells differentiate into memory cells and others in to antibody-secreting plasma cells. (B) The Pre-BCR consists of a rearranged Ig H chain and the SLC chain, which has two invariant polypeptides λ5 and V_{preB}

Figure 2

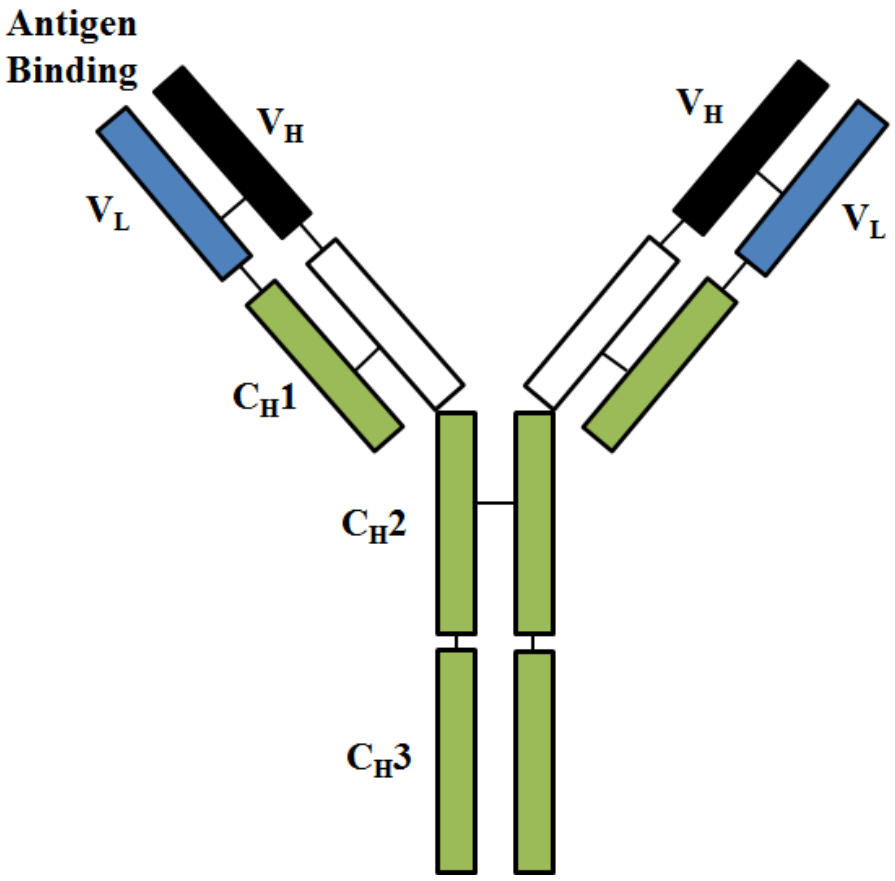


Figure 2. Structure of Antibody Molecule. Antibody molecules are composed of two identical heavy chains (VH) and light chains (VL). Each VH and VL has an antigen-binding variable region that determines the specificity of the antibody and a constant region that communicates the effector function. All the chains in an antibody are held together by disulfide bonds. The variable region can be further divided into four framework regions and three complementarity determining regions. Some heavy chain constant regions have four CH domains, while others have three.

Figure 3

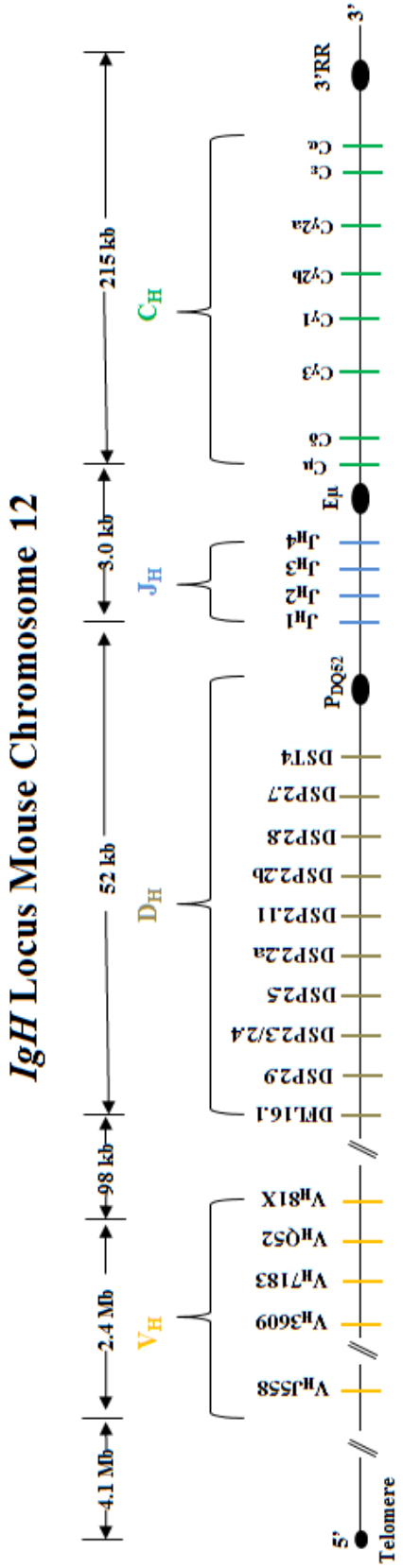

IgH Locus Mouse Chromosome 12
5'
Telomere
3'
3'RR
Eμ
PDQ52
VH
DH
JH
CH
VH81X
VHQ52
VH7183
VH3609
VHJ558
DFL16.1
DSP2.9
DSP2.3/2.4
DSP2.5
DSP2.2a
DSP2.11
DSP2.2b
DSP2.8
DSP2.7
DST4
JH1
JH2
JH3
JH4
Cμ
Cδ
Cγ3
Cγ1
Cγ2b
Cγ2a
Cε
Cα
4.1 Mb
2.4 Mb
98 kb
52 kb
3.0 kb
215 kb

Figure 3. Organization of the 3 Mb *IgH* Locus in Germline Configuration, Mouse Chromosome 12. Schematic illustration above is based on C57BL/6 (Johnson et al., 2006), F1 long arm of mouse chromosome 12. The names of each of the four gene segment families is above each respective region. Also included are the three major *cis*-regulatory elements of the *IgH* locus P_{DQ52} , Eμ, and the 3'RR. The most 5' variable gene is 4.1 Mb from the telomere. Figure is not drawn to scale.

Figure 4

A

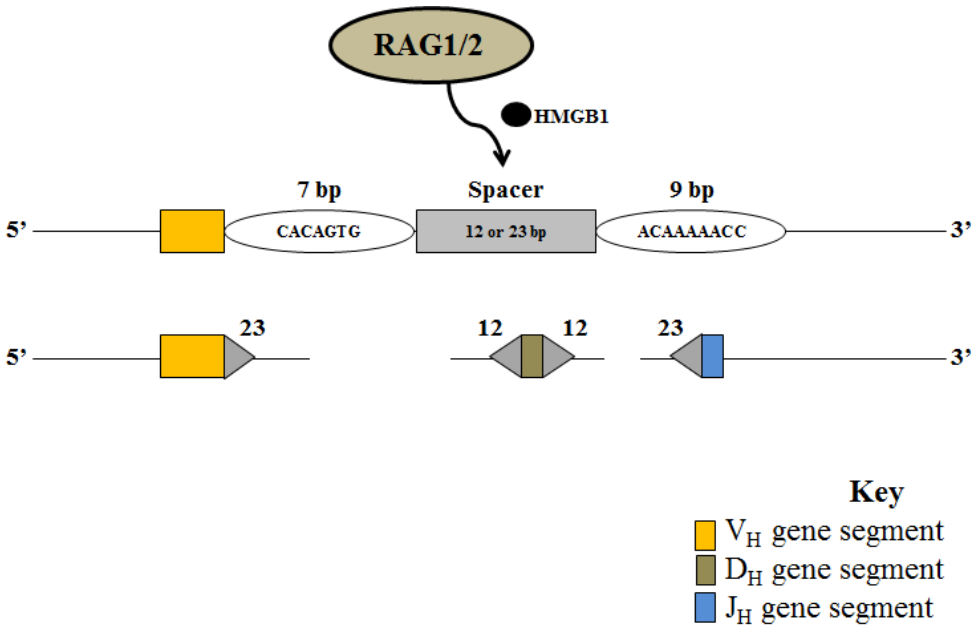

RAG1/2
HMGB1
7 bp
Spacer
9 bp
5'
CACAGTG
12 or 23 bp
ACAAAAACC
3'
23
12
12
23
5'
3'
Key
V_H gene segment
D_H gene segment
J_H gene segment

Figure 4

B

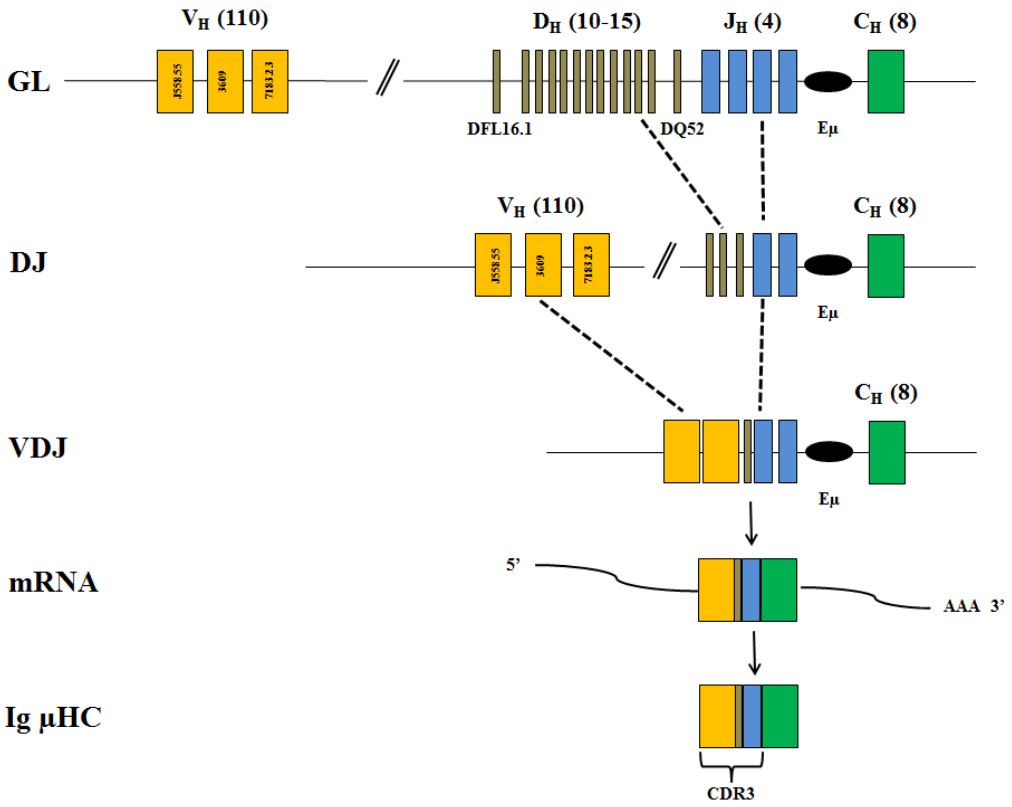

GL
DJ
VDJ
mRNA
Ig μHC
V_H (110)
D_H (10-15)
J_H (4)
C_H (8)
J558.55
3609
7183.2.3
DFL16.1
DQ52
Eμ
5'
AAA 3'
CDR3

Figure 4. RAG and RSS engagement. (A) Depicted above is the recombination signal sequence (RSS) that has a heptamer, a spacer of 12 or 23 bp, and a nonamer. Only a 12-bp spacer can recombine with a 23-bp spacer gene segment. These RSS sequences are adjacent to each antigen receptor gene segment. In the initial step of gene rearrangement, RAG 1 and 2, bind to either the 12-bp RSS or 23-bp RSS, in the presence of HMGB1. (B) The rearrangement process shown above is that for the *IgH* locus The D (diversity) and J (joining) gene segments undergo rearrangement first, followed by rearrangement of the V (variable) gene segment to the DJ locus. The complete rearranged V(D)J locus is first transcribed into mRNA and then translated into the μ-heavy chains. Shown at bottom is the hypervariable region of the antigen receptors, known as CDR3 and encoded by the rearranged V(D)J gene segment.

Figure 5

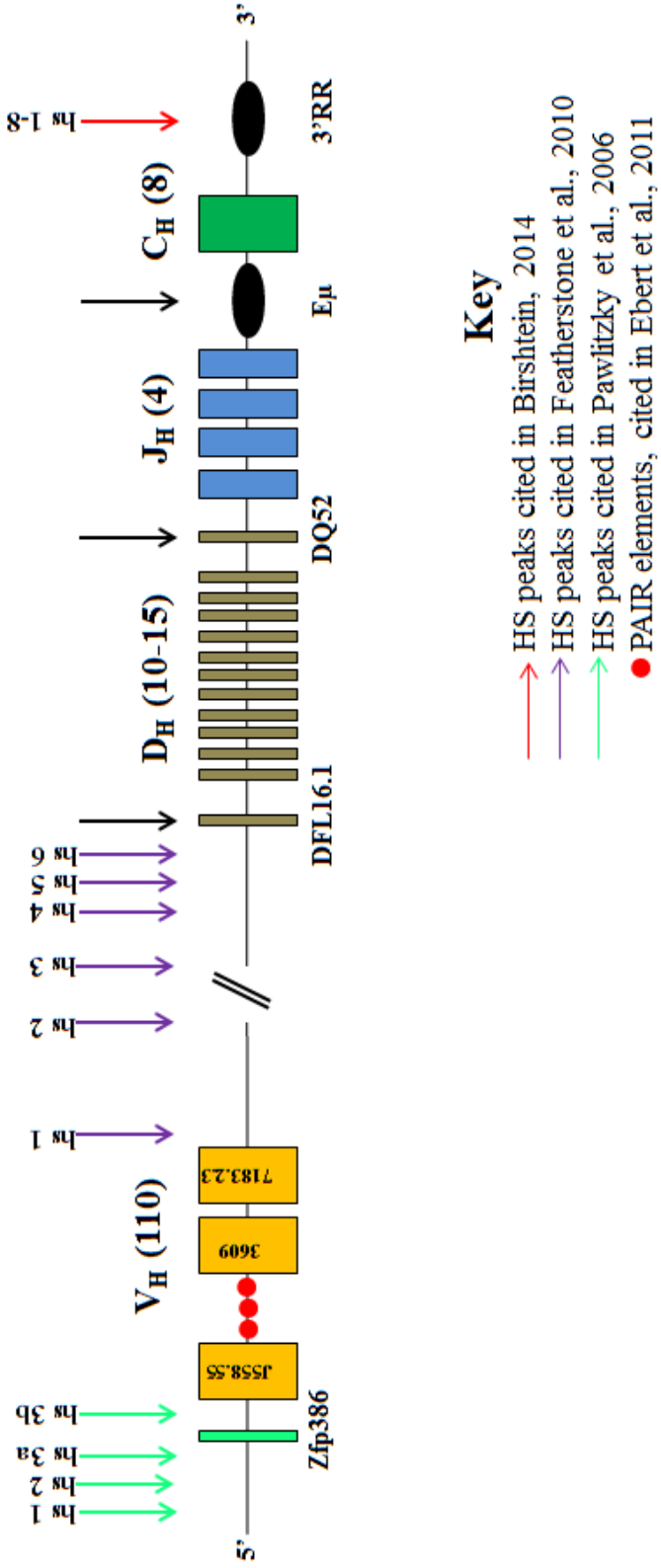

3'
3'RR
hs 1-8
C_H (8)
Eμ
J_H (4)
DQ52
D_H (10-15)
DFL16.1
hs 6
hs 5
hs 4
hs 3
hs 2
hs 1
V_H (110)
7183.23
3609
J558.55
Zfp386
hs 3b
hs 3a
hs 2
hs 1
5'
Key
HS peaks cited in Birshtein, 2014
HS peaks cited in Featherstone et al., 2010
HS peaks cited in Pawlitzky et al., 2006
PAIR elements, cited in Ebert et al., 2011

Figure 5. Overview of *cis*-acting elements on the *IgH* locus. This schematic illustration gives an overview of the *cis*-regulatory elements on the *IgH* locus that have been reported. It does not include CTCF elements that have been reported throughout the IgH locus. Shown are the DNase I hypersensitivity peaks that have led to *cis*-regulatory additional studies to determine the function of these sites. While some have had a significant role in *cis*-regulation of the *IgH* locus (Eμ, 3'RR), others have no functional effect (Pawlitzky et al., 2006).

Figure 6

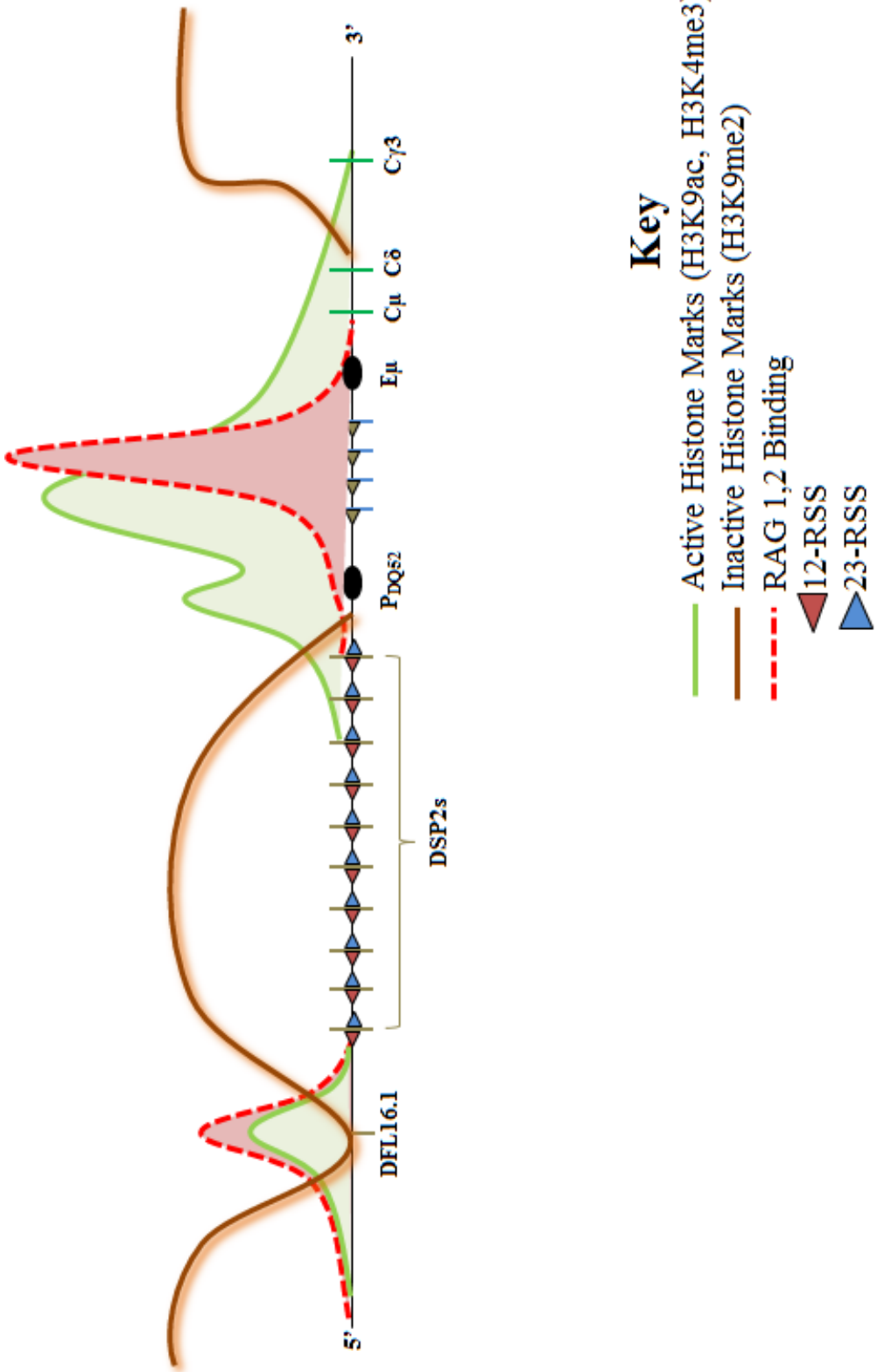

3'
Cγ3
Cδ
Cμ
Eμ
PDQ52
DSP2s
DFL16.1
5'
Key
Active Histone Marks (H3K9ac, H3K4me3)
Inactive Histone Marks (H3K9me2)
RAG 1,2 Binding
12-RSS
23-RSS

Figure 6. RAG Association and Histone Modifications on Germline *IgH* Locus. This schematic is updated from Sen and Oltz 2006 to represent the pattern of histone modifications and RAG association as currently known on mouse chromosome 12, germline configuration. Active histone marks (green) closely follow RAG binding (red), and suppressive histone marks (brown) dominate the DSP region. RSS segments are depicted to demonstrate accessibility as it's related to histone modification and RAG binding.

Chapter 2

Eμ Enhancer-Induced Changes In Chromatin Structure & Function

Introduction

Gene regulation in mammals is central to understanding cellular development. Enhancers are critical *cis*-regulatory elements that recruit transcription factors to modify chromatin structure for coordinated gene expression. Initially discovered in the genome of simian virus 40 (SV40), enhancers have been observed in transient expression assays to activate transcription independent of their distance relative to their gene promoters (Banerji et al. 1981; Ong et al., 2011). The Eμ intronic enhancer was the first transcriptional tissue-specific enhancer discovered, as it was able to drive the expression of immunoglobulin or other transgenes in a position independent manner in lymphoid cells alone (Alt et al., 1982; Banerji et al. 1983; Gilles et al., 1983; Neuberger 1983). As a *cis*-regulatory element, the Eμ enhancer has been demonstrated to have a critical role in both the expression of genes and recombination events at the pro-B cell stage of mouse B-cell development.

Located on mouse chromosome 12, the Eμ intronic enhancer is part of the 3 Mb immunoglobulin heavy (*IgH*) chain gene locus that encodes antigen receptors for lymphocytes (Chowdhury and Sen, 2004). The *IgH* locus contains variable (V_H), diversity (D_H), joining (J_H) gene segments and constant (C) region exons. These V(D)J gene segments participate in two recombination events to produce a complete rearranged *IgH* allele in pre-B cells. In a precise and stepwise order, gene rearrangement between the D to J_H loci is always initiated in pro-B cells prior to the V to DJ_H rearrangement. Both "recombination centers" and the "accessibility hypothesis" explain this ordered sequence of recombination events (Ji et al., 2010; Yancopoulos and Alt, 1985).

First, in order for gene rearrangements to proceed on the *IgH* locus, RAG recombinase must be recruited to the antigen receptor genes to introduce double strand breaks at accessible recombination signal sequences (RSSs). The recombination center is the formation of local RAG-binding regions that serve as the selected sites where antigen receptor gene segments are captured for recombination. The highest density of RAG binding in pro-B cells has been identified in the J_H locus, which partly explains why the initial recombination step is between the D_H and J_H gene segments (Ji et al., 2010).

Second, the accessibility hypothesis originally suggested that it was the presence of accessibility control elements (ACE) at the individual antigen receptor loci that lead to ordered genetic rearrangements (Yancopoulos and Alt, 1985). Using a combination of transgenic models and *in vitro* systems, both transcriptional enhancers and promoters were found to be *cis*-acting ACEs that regulate the cell type-specific chromatin structure to determine the targeting of RAG recombinase. It is then precisely the coordinated accessibility of both an antigen receptor gene locus and its ACEs to the RAG recombinase that determines the specific selection of D_H and J_H antigen receptor genes to initiate recombination in pro-B cells. Such *cis*-regulatory ACEs are then central to the core features of both the accessibility hypothesis and the local RAG-bound recombination centers.

One significant *cis*-acting ACE that has been identified to have a significant role related to D_H to J_H recombination in pro-B cells is the 700 bp Eμ enhancer. Situated between the J_H-Cμ intron, deletion of the Eμ enhancer led to a significant reduction in D_H to J_H rearrangements and a major defect in the subsequent V_H to DJ_H rearrangement (Afshar et al., 2006; Perlot et al., 2005). The Eμ enhancer has a 220 bp core region (cEμ)

and is surrounded on each side by matrix attachment regions (MARs) (Gillies et al., 1983; Jenuwein et al., 1997). Deletion of cEμ resulted in reduction in transcription and transcription-associated histone modifications, as well as significant impairment of D_H to J_H gene rearrangement (Chakraborty et al., 2009). These investigations into both Eμ and cEμ have demonstrated the enhancer's pivotal role as a *cis*-regulatory element that controls epigenetic modifications, transcriptional regulation, and D_H to J_H recombination

Given the historical view of enhancers as able to activate transcription independent of their distance relative to the promoters of gene, this project directly tested whether or not the location of the Eμ intronic enhancer in the *IgH* locus is required for its gene regulatory function and if the enhancer's function as it relates to gene rearrangement changes based on the location. No studies have examined within an *in vivo* setting whether the distance of an enhancer is critical for the appropriate gene expression and resulting phenotype. For this investigation, we generated bacterial artificial chromosome (BAC) transgenic mice that contain the Eμ intronic enhancer in different locations and assayed chromatin, transcription, and recombination. Using this system, we were able to test the hypothesis that enhancers are *cis*-acting elements that activate transcription regardless of location to their gene promoters. Our results demonstrate that regions of heterochromatin can be changed to euchromatin from the location of the Eμ intronic enhancer, and that when the location of the Eμ intronic enhancer is changed, unique transcripts and rearrangements are induced compared to the endogenous state. Therefore, the precise location of Eμ is essential to generate normal chromatin structure, optimal transcription and D_H to J_H rearrangements. From this study of the Eμ enhancer in

different locations on modified BACs, a new view of enhancers has emerged where their unique location in a locus is required for normal gene regulation.

Experimental Flow

BAC transgenic mice were developed as described in the Materials and Methods section. A schematic representation of the 11 kb BAC construct and 181 kb insert is given in Figure 7, and an overview of the modified BACs with the relocated Eμ enhancer is in Figure 8. In brief, C57BL/6 mice carrying the modified BAC constructs were generated by recombineering experiments carried out by Dr. Changying Guo and the injections of the modified BACs were performed at the University of Connecticut. Transgene positive C57BL/6 founder mice were initially bred to J_HT mice that have a 4 kb deletion between the Xho I and EcoR I sites that are directly upstream of DQ52 and downstream of Eμ, respectively (Gu et al., 1993). The purpose of this breeding was to ensure that all measurements reflected the transgene and not endogenous *IgH* alleles. After this initial breeding, transgene positive mice were subsequently bred to J_HT. Founder mice of different transgenic strains were separated by mouse tags in the BRC Mouse Core Facility. Each transgenic strain had two sets of founders except EU+. To test for the presence of the transgene in the mice prior to experiment, genotyping was performed using several sets of primers that anneal to BAC specific regions that are outlined in the Material and Methods section and shown in Figure 9. In all experiments 6 mice, ages 8-10 weeks, were sacrificed and bone marrows were flushed for each individual transgenic strain. Cells were collected and $CD19^+$ B cells were isolated by positive selection using Stem Cell's Robosep Isolation System (Stem Cell #18754). Approximately 5 million cells were present after isolation, and from this pool, three divisions were made for the cells: (a) 5×10^5 cells were used for subsequent RNA isolation, (b) 5×10^5 were used for DNA isolation for downstream rearrangement assays

and (c) the remaining cells were plated for expansion onto an OP-9 stromal layer as described in the Material and Methods. Both FACs analysis and Southern Blots were performed to confirm pro-B cell identity and transgene presence on J_HT background in experimental mice (Figures 10 and 11). All probes are listed in the Material and Methods section. Southern blots also measured transgene copy number, which was quantified for both founder sets of mice using Qiagen's Biomarker Copy Number PCR Assays (Catalog #, VPM112-0573154A) (Figure 12).

Results and Discussion

Modified BAC Transgenic Overview

In order to test the hypothesis that enhancers are *cis*-regulatory sequences that activate transcription regardless of location to their gene promoter, six different transgene constructs from the RP23-458I14 BAC were modified based on a 181 kb sequence from mouse chromosome 12 (Figure 7a). Each modified BAC represented the 181 kb region between positions 114,567,000 – 114,748,000 on the C57BL/6 mouse chromosome 12 (Figure 7b). This region starts 80 kb upstream of DFL16.1 (chromosome 12: 114,668,722) and ends approximately 5 kb downstream of Cγ1 (Figures 3, 7B). Importantly, there are no V gene segments in the BAC (Figure 7c). This allowed the study to focus solely on the effects related to the D_H and J_H gene segments with respect to the relocation of the Eμ enhancer in the modified BACs.

The constructs were specifically designed to test the central hypothesis concerning whether the Eμ intronic enhancer has the same gene regulatory functions on the *IgH* locus independent of distance from its endogenous site. A schematic representation of each BAC construct is depicted in Figure 8. As illustrated in Figure 12, the BACs were given the following labels: EU+, EU-, DSP, DPEU, VD, and VDEU. The first pair, EU+ and EU-, are constructs identical to the aforementioned defined area of C57BL/6 mouse chromosome 12, with the exception that the EU- construct has the endogenous 700 bp Eμ deleted that is between the Xba I and EcoR I sites (chr12: 114,665,495 – 114,666,183). These two BAC constructs allow comparison to prior chromatin, transcription, and recombination experiments performed in pro-B cells in

order to address whether the transgenes were able to recapitulate the gene regulatory landscape in pro-B cells (Chakraborty et al., 2009).

The next pair of BACs, DSP and DPEU (Figure 8) represented the first modified BACs where the Eμ enhancer was relocated. These BACs were identical in that each contains the Eμ enhancer in the middle of the characterized heterochromatic D_H-region on mouse chromosome 12 at position 114,697,256 - 114,697,956, which is 31 kb upstream from the endogenous site. This region is flanked by DSP2.x upstream and DSP2.3/4 downstream. However, there was one critical difference between the DSP and DPEU transgenes. In the case of DSP, there is no Eμ enhancer in the endogenous location, whereas the DPEU construct contains two Eμ enhancers, one in the endogenous location and one within the D_H - gene segments. This allowed any effect observed from the relocated Eμ enhancer to be associated with the Eμ enhancer and no other gene regulatory element.

To test whether a distant Eμ enhancer was able to regulate gene expression in the same way as the endogenous Eμ enhancer, we designed the VD and VDEU constructs. These constructs had the Eμ enhancer located 5 kb upstream of DFL16.1 (chromosome 12:114,720,388) and 59 kb from the endogenous location. While both the VD and VDEU constructs contained the relocated Eμ, only the VDEU construct had both the relocated Eμ enhancer and the Eμ enhancer in the endogenous location. This allowed any effect detected from the distant relocated Eμ enhancer to be associated with the Eμ enhancer and no other *cis*-acting element. In summary, the DSP transgene allowed the study of both distance and whether regions of heterochromatin are able to be changed by the presence of an enhancer. The VD construct allowed the question of distance to be

answered, as the Eμ enhancer is 59 kb from the endogenous site. Together, these constructs and the outlined control BACs, EU+ and EU-, were able to fully examine whether enhancers can regulate gene expression independent of location. Each of these modified BAC inserts were placed into the pBACe3.6 vector and delivered into C57BL/6 mice by pronuclear injection.

Epigenetic and Transcription Profiles of EU+ and EU- Transgenes

In order to compare the modified transgenic mice, we set out to determine whether the control set of transgenes, EU+ and EU-, recapitulated the transcriptional and epigenetic profiles of chromosomal *IgH* alleles (Chakraborty et al., 2009). For epigenetic analysis, we carried out chromatin immunoprecipitation studies from expanded OP-9 pro-B cells using antibodies directed against histone modifications associated with gene activation (H3K9ac and H3K4me3) and gene inactivation (H3K9me2). With respect to active chromatin modifications in EU+, both H3K9ac and H3K4me3 had a distinguishable pattern where the greatest presence of these active marks was at J_H2 (Figure 13A, B, blue bars). In EU+, there was a stepwise pattern that started with low active marks at Cγ3 and peaks at J_H2. There was then a subsequent decrease in active marks at DQ52, and this was reduced further as minimal active histone marks were detected upstream of DQ52 in the EU+ mice. This epigenetic profile was very similar to that of endogenous *IgH* alleles (Chakraborty et al., 2009).

In the EU- transgene, both H3K9Ac and H3K4me3 levels were reduced at Cμ compared to EU+ alleles (Figure 13A, B, red bars). Outside of these two sites, active marks were not detected in the EU- transgene. The pattern of H3K9me2, a marker of

inactive chromatin, was similar in both EU+ and EU- transgenes, but the EU- transgene had elevated H3K9me2 at both J_H2 and DQ52 (Figure 13c, blue and red bars, respectively). From Cμ to DFL16.1, there was a successive increase in H3K9me2 with significant elevation in the D_H region. Just as the D_H region is dominated by heterochromatin marks in pro-B cells, the EU+ and EU- transgenes shared the same heterochromatin pattern. The DSP gene segments have high levels of heterochromatin regardless of whether Eμ is present or absent. Putting both the active and inactive chromatin patterns together, the EU+ and EU- BAC transgenes effectively represent the gene regulatory pattern of Eμ with respect to active and inactive chromatin modifications as found in the endogenous *IgH* locus of mouse pro-B cell lines. (Chakraborty et al., 2009) (Figure 6).

Since the Eμ enhancer in the EU+ control transgene had a similar epigenetic profile to that in Eμ sufficient mouse pro-B cell lines, we determined the Eμ enhancer in the EU+ control transgenic mice was also able to induce transcription of genes due to the permissible chromatin structure at different loci. In the EU+ control transgene, Eμ dependent transcription of genes was observed only at DQ52 when compared to the EU- transgene (Figure 14A, blue and red bars, respectively). Upstream of DQ52, there is minimal transcription throughout the D_H domain in both the EU+ and EU- transgenes which reflects the heterochromatin structure of the DSP region. Therefore, both the chromatin and transcription profiles in the EU+ and EU- BAC transgenes effectively reflected the gene expression pattern of Eμ as found in the endogenous locus of mouse pro-B cells (Chakraborty et al., 2009) (Figure 6). We were then able to test the gene

regulatory effects of the relocated Eμ enhancer in the four other constructs: DSP, DPEU, VD, and VDEU.

Epigenetic and Transcription Profiles of Relocated Eμ enhancer in DSP Region

After observing close similarity between the regulation patterns of the control BAC transgenes and those of the endogenous *IgH* alleles, we investigated the effects of relocating the Eμ enhancer into the heterochromatic DSP region. We found that both activation marks (H3K9ac, H3K4me3) were present in the DSP region when the Eμ enhancer was relocated into the DSP gene segments in the DSP transgenic mouse (Figure 13A and B, green bars). The pattern between DSP and EU- was similar at Cγ3, Cμ, and Eμ+1, which is attributed to both transgenes having a deleted Eμ from the endogenous location. At J_H2, a significant increase in active histone marks was observed in DSP compared to EU- transgenic alleles, however, levels did not reach to those present in EU+ alleles. This active region peaked at DQ52 and then decreased to DFL16.1, though it was still detectable and was nearly ~2.5 times higher than EU-. Conversely, inactive H3K9me2 marks for DSP were significantly reduced throughout the D_H region from the relocated Eμ (Figure 13C). Therefore, with a relocated Eμ enhancer in the middle of the DSP gene locus, chromatin marks shift from a region dominated with heterochromatin in EU+ and EU- settings, to active marks that are associated with gene activation in the DSP transgenic mouse.

The pattern for double-enhancer alleles in DPEU transgenic mice was similar to DSP in the D_H region but at higher amounts for H3K9ac and H3K4me3 (Figure 13D and 13E, purple bars). In addition, for the DPEU transgene, the Eμ enhancer located in the

endogenous location followed the same pattern for active and inactive marks compared to the EU+ transgene between Cγ3 and J_H2. Consequently, the Eμ enhancer in the endogenous setting for the DPEU transgene was able to maintain active chromatin marks within a set distance that was distinguished from the DSP transgene. This was confirmed in DSP, as the relocated Eμ enhancer was able to induce chromatin changes within the D_H gene segment only, and not downstream of J_H2. We conclude that the Eμ enhancer is able to 'open' heterochromatin and induce active chromatin marks.

The presence of new euchromatin in the DSP transgene was reflected in unique transcripts that were induced at DSP2.3 and remained throughout the D_H domain (Figure 14A, green bars). A similar pattern, but elevated transcript levels were observed for DPEU (Figure 14A, purple bars). Therefore, without the Eμ enhancer, the chromatin in the D_H region is closed and the genes are transcriptionally silent. When the Eμ enhancer is relocated to the middle of the D_H region, the chromatin configuration changes and unique transcripts are induced in nearby genes.

Epigenetic and Transcription Profiles of Relocated Eμ enhancer 5' DFL16.1

In the last level of comparison, when the Eμ enhancer was located 59 kb from the endogenous location in the VD transgenic model, only local chromatin remodeling for active histone marks was observed at DFL16.1 (Figure 13G, H, black bars). In every other region analyzed, the VD transgene was similar to the EU- transgene for both active and inactive histone modifications (Figures 13G - I). While VDEU has active histone marks between Cμ and J_H2, these modifications were attributed to the Eμ enhancer in the endogenous location and not the distal Eμ enhancer (Figure 13G, H, orange bars).

Similar to DPEU, the VDEU transgene was also able to restore and maintain the chromatin structure in the endogenous Eμ, peaking at J_H2. Therefore, in the distant Eμ enhancer VD transgene, the effects of the relocated Eμ enhancer were isolated to DFL16.1 and the Eμ enhancer was unable to affect chromatin structure from a distant location.

Similar to the chromatin pattern for the VD transgene, only minimal transcripts were observed throughout the D_H domain, with a slight increase in transcript level at DFL16.1 (Figure 14B, black bars). The VDEU transgene was similar, but had significant transcription at DQ52 and this was attributed to the Eμ enhancer located at the endogenous location (Figure 14B, orange bars). Therefore, the main theme that emerged was that Eμ enhancer was not able to generate the same chromatin or transcription patterns independent of location, but was able to affect unique chromatin and transcription patterns that were dependent upon the enhancer's location.

D_H to J_H Rearrangement in Transgenic Mice

Next, we set out to determine if relocating the Eμ enhancer had any functional effects with respect to gene recombination. In order for gene rearrangement to take place between the D_H and J_H gene segments, an RSS must be accessible at each unrearranged gene so that RAG recombinase may be recruited to the loci targeted for rearrangement. Each RSS sequence consists of a conserved heptamer and nonamer. These sequences are separated by either a 12 or 23 bp spacer sequence that is less conserved and these spacers define the two types of RSSs, 12-RSS and 23-RSS. Known as the 12/23 rule, efficient recombination only takes place between a 12-RSS and a 23-RSS.

In order to study gene rearrangement in the transgenic mice, genomic DNA was amplified from isolated $CD19^+$ pro-B cells using 5' primers that hybridize to either DFL16.1 or multiple DSP gene segments, and a common 3' primer that is located downstream of J_H3. We used the non-rearranged β_2-microglobulin gene (β_2M) to normalize DNA used for the recombination assays. As shown in Figure 15, comparable levels of gene recombination were detected in the EU+ strain when compared to C57BL/6 bone marrow cells isolated by CD19+ selection. As expected, minimal gene rearrangements were detected in the EU- transgenic mice. Therefore, the EU+ transgenic allele was able to effectively induce gene rearrangements. We consistently detected rearrangements using both the DFL16.1 or DSP primers in DSP mice that were either comparable to EU+ or slightly less. When the relative band signal intensity was quantified using the Gene Tools Syngene program for the DSP to J_H gene rearrangements, the DSP transgene had a greater intensity than EU+ for DSP-J_H2 and DSP-J_H3 rearranged alleles, but less than DPEU for each rearrangement (Figure 15B). For the DPEU mouse strain, gene rearrangements were of greater intensity than DSP due to having both the relocated Eμ and the endogenous Eμ. In the VD strain, few rearrangements were detected using either the DSP or DFL16.1 primer, which were comparable to the EU- strain (Figure 15). For the VDEU strain, all gene rearrangements were comparable to EU+, and thus the Eμ in the endogenous location and not the distal Eμ enhancer was directing gene recombination events on this allele (Figure 15). Therefore, while the Eμ enhancer was able to direct gene rearrangements, it was only the single Eμ enhancer in the DSP strain that was able to induce unique recombination events at a level comparable to the endogenous setting using either DFL16.1 or DSP primers.

This was reflected in both the chromatin and transcription data, where the typically heterochromatic DSP locus was opened in the DSP mouse and had both active histone marks and unique transcripts. These characteristics then allowed the Eμ enhancer in the middle of the DSP region to activate gene recombination events with any of the associated J_H segments. As the chromatin and transcription data indicated minimal locus opening from the distal Eμ enhancer in the VD mouse, it was also observed to have minimal gene rearrangements and those detectable rearrangements were from the DFL16.1 primer that is located 5 kb near the relocated Eμ enhancer. This data indicates that Eμ can activate D_H recombination only from select locations within the D_H - Cμ region.

Next, we set out to determine whether the relocated Eμ enhancer affected all DSP gene segments equivalently. Since the chromatin structure was opened by the presence of the Eμ enhancer in the middle of the D_H gene segments, it was critical whether all of the RSS sequences from the D-gene segments were used for gene rearrangement as a result of the chromatin opening, or whether the Eμ enhancer was only able to affect some D_H gene segments over others. In order to determine the D-gene usage in these recombination events in the EU+, DSP, and DPEU transgenic mice, 90 PCR amplicons from the DJ_H rearrangement gels were cloned and sequenced (Figure 16). Due to the heterochromatin structure of the D_H locus in the EU+ transgene, it was predicted that the Eμ enhancer in the EU+ transgene would initiate rearrangements with one of the outer and accessible genes in higher amounts than the interior, inaccessible genes. In the EU+ strain, such rearrangements were observed to be mainly at the DSP2.2 gene segment, but then dispersed throughout all the other DSP gene segments at much lower levels (Figure

16). In contrast to the pattern of rearrangements in EU+, we found that for the relocated Eμ enhancer in the DSP and DPEU strains, the majority of gene rearrangements occur at DSP2.3/4, which is the first downstream DSP gene of the relocated Eμ enhancer. This was evident for both the DSP and DPEU transgenes. Although there are active histone proteins throughout the D_H gene locus, the DSP2.3/4 was used in much greater proportion than any other gene segment. We attribute the selection of this gene segment to the Eμ enhancer affecting the closest gene promoter to induce gene rearrangements, with relatively little effect on other DSP gene segment.

RAG Recombinase Accessibility in Transgenic Mice

Next, we set out to determine the mechanism of this functional recombination data. It has been previously demonstrated that binding of RAG protein to accessible DNA takes place in a focal manner, and is limited to sites that have elevated levels of H3K4me3 and RSS sequences. The stationing of these local RAG-bound areas are known as "recombination centers." Such regions of distinction are developmental and lineage-stage specific and their purpose is to coordinate V(D)J recombination so as to provide the specific sites where the antigen receptor gene segments are captured for recombination (Ji et al., 2010).

To probe for the presence of recombination centers in the different transgenes, and to see whether or not the RAG protein was driving these recombination events, we performed RAG ChIP experiments to detect the presence of RAG across different D_H and J_H gene segments. We observed that the presence of RAG1 and RAG2 in the J_H gene segments was highest in each strain that had both Eμ enhancers (EU+, DPEU, VDEU),

and only in the DSP strain was this focal establishment of RAG-1/2 in the J_H region maintained (Figure 17). Minimal levels of RAG1 and RAG2 protein were detected in both the EU- and VD strains. As one moves upstream from the J_H genes and into the D_H region, the level of RAG binding at DQ52 drops off, and decreases further at DSP2.2 in EU+, DSP, DPEU, and VDEU. It then increases at DFL16.1 in each of these strains to levels that are above DQ52 yet below both J_H2 and J_H4. As a result, recombinase accessibility to the recombination center remains highest in the J_H gene segments for those strains that have the Eμ in the endogenous location (EU+, DPEU, and VDEU), and only the single relocated Eμ enhancer in the DSP mouse strain demonstrates a similar pattern (Figure 17). While the Eμ enhancer in the DSP mouse is able to induce unique rearrangements, it is because of the focal binding of RAG-1/2 to the J_H gene segments and the presence of the recombination center that the Eμ enhancer in the D_H locus is able to induce gene rearrangement events. When this focal binding of RAG is removed, such as the VD mouse, minimal rearrangements are able to take place as no recombination center is present.

In conclusion, this project demonstrated that the Eμ enhancer is able to induce unique transcripts and recombination events in the DSP mouse, as it is able to open the local heterochromatin structure. This configuration allows the focal establishment of RAG-1/2 binding in the J_H gene segments and the existence of the recombination center to remain, so that unique recombination events occur as a result of the relocated Eμ enhancer. However, these observations are limited to the DSP mouse where the Eμ enhancer is 31 kb upstream from the endogenous site and not the VD mouse where the Eμ enhancer is located 59 kb upstream from the endogenous site. In the VD transgenic

lines, transcription and active chromatin modifications were limited to DFL16.1, though at lower levels. Therefore, the VD mouse essentially represented the EU- mouse with the only difference of the enhancer's location 5 kb upstream of DFL16.1. We conclude that the recombination between D_H and J_H gene segments is dependent on the location of the Eμ enhancer. In light of this, the *in-vitro* definition that enhancers are *cis*-regulatory sequences that activate transcription regardless of location and orientation to their gene promoters must be reconsidered at least for the Eμ enhancer. This hypothesis is not true, as the Eμ enhancer's precise location is essential to generate normal chromatin structure, optimal transcription, D_H to J_H rearrangements, and pro-B cell development. In this study, chromatin structure dictated whether or not transcription and recombination can occur, and this was dependent on the extent to which the Eμ enhancer was able to open the chromatin structure. In the DSP mouse strain, the Eμ enhancer changed the local chromatin configuration to histone marks associated with gene activation, while the extent was very minimal in the VD mouse strain. Thus, not only was the chromatin structure changed in the DSP mouse from the Eμ enhancer, but the open chromatin structure configuration allowed the RAG-1/2 recombination center to be established, leading to gene recombination events that were unique to the relocated Eμ enhancer. We conclude the precise location of the Eμ enhancer is required for normal chromatin configurations, optimal transcription, D_H to J_H rearrangements, and pro-B cell development.

Figure 7

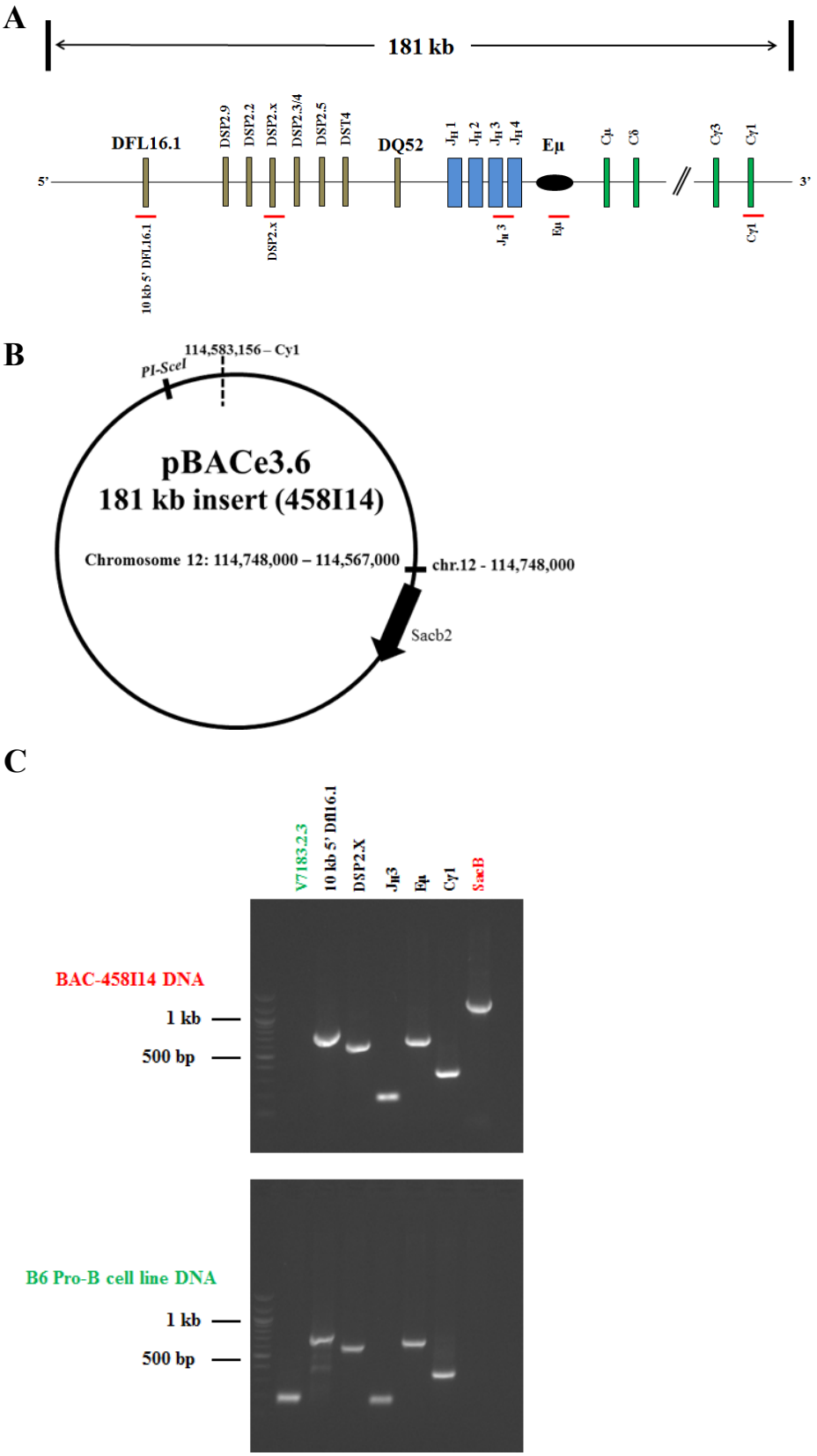

A
181 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
10 kb 5' DFL16.1
DSP2.x
JH3
Eμ
Cγ1
B
114,583,156 – Cγ1
PI-SceI
pBACe3.6
181 kb insert (458I14)
Chromosome 12: 114,748,000 – 114,567,000
chr.12 - 114,748,000
Sacb2
C
V7183.2.3
10 kb 5' Dfl16.1
DSP2.X
JH3
Eμ
Cγ1
SacB
BAC-458I14 DNA
1 kb
500 bp
B6 Pro-B cell line DNA
1 kb
500 bp

Figure 7. Schematic representation of BAC construct and insert. (A) Each modified BAC originated from RP23-458I14 and contains the region between ~~1~~14,748,000 – 114,567,000 on mouse chromosome 12 as represented by the mm9 sequence assembly, http://genome.ucsc.edu. (B) The BAC vector sequence is 11 kb and contains a PI-SceI linearization site, along with the Sacb selection gene. (C) The BAC constructs contained no VH -gene segments and was confirmed to be 181 kb by PCR. Shown on the top gel are different sets of primers that were used to map the BAC. Included is the SacB primer, which is specific for the backbone of the BAC vector. The bottom gel is DNA from the B6D345 pro-B cell line that does not contain the BAC, and therefore, this cell line has VH -genes, but no SacB.

Figure 8

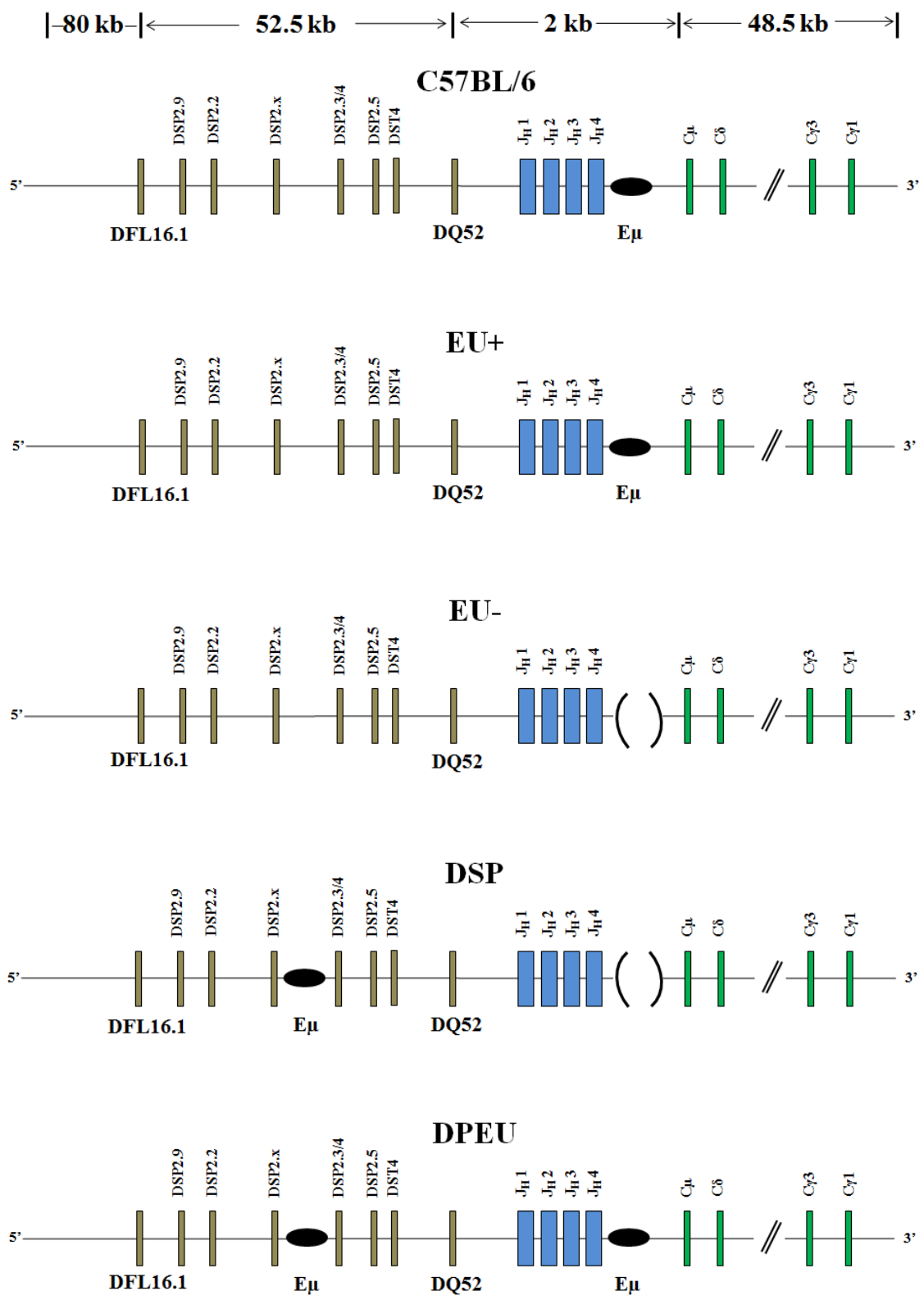

|–80 kb–|
52.5 kb
2 kb
48.5 kb
C57BL/6
EU+
EU-
DSP
DPEU
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
JH1
JH2
JH3
JH4
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1
DQ52
Eμ

Figure 8 (cont'd)

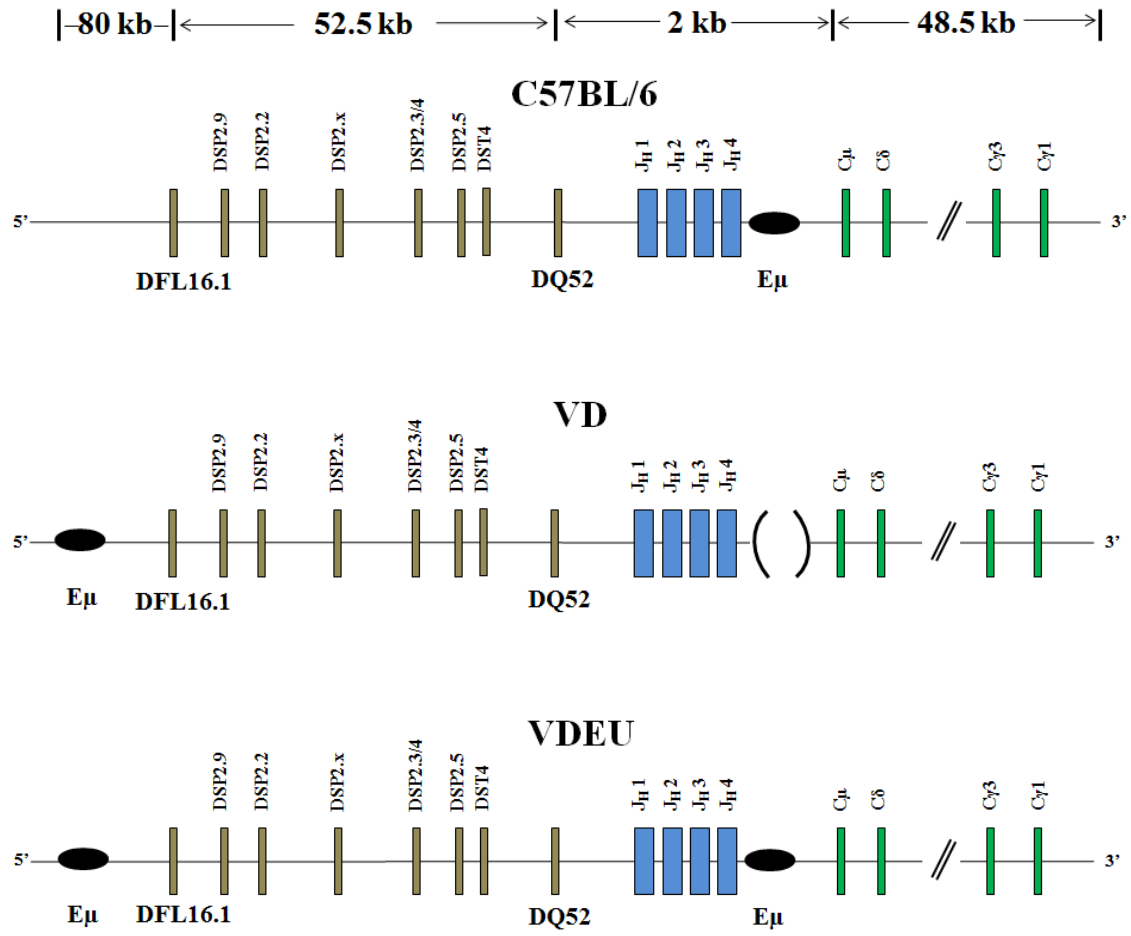
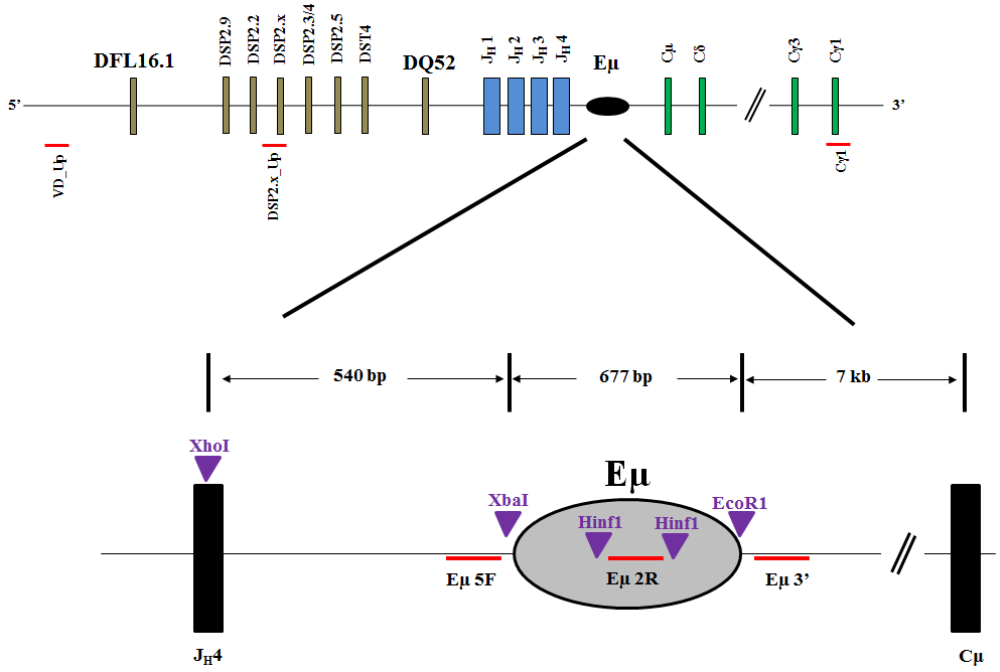


Figure 8. Modified BAC construct overview. The endogenous *IgH* locus of C57BL/6 mice is depicted at the top of the diagram. Underneath are the six modified BAC constructs that reflect either the Eμ enhancer in the endogenous location (EU+), deleted location (EU-), relocated positions (DSP, VD), or in both the endogenous and relocated positions (DPEU, VDEU). These relocated positions were selected based on distance and chromatin structure. The Eμ enhancer in the VD mouse is located 57 kb away from the endogenous site, while the Eμ enhancer in the DSP mouse is located 31 kb and in area of characterized in the endogenous IgH locus as one of heterochromatin. All constructs originated from the RP23-458I14 BAC.

Figure 9

A


DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
Jμ1
Jμ2
Jμ3
Jμ4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
VD_Up
DSP2.x_Up
Cγ1
540 bp
677 bp
7 kb
XhoI
Eμ
XbaI
Hinf1
Hinf1
EcoR1
Eμ 5F
Eμ 2R
Eμ 3'
JH4
Cμ

B

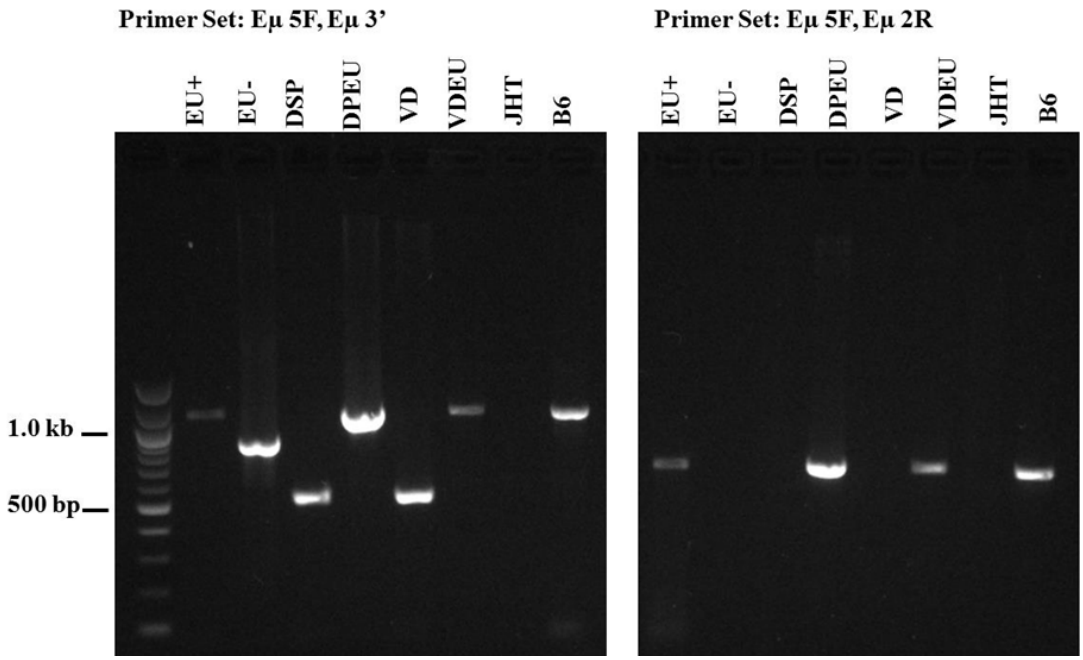
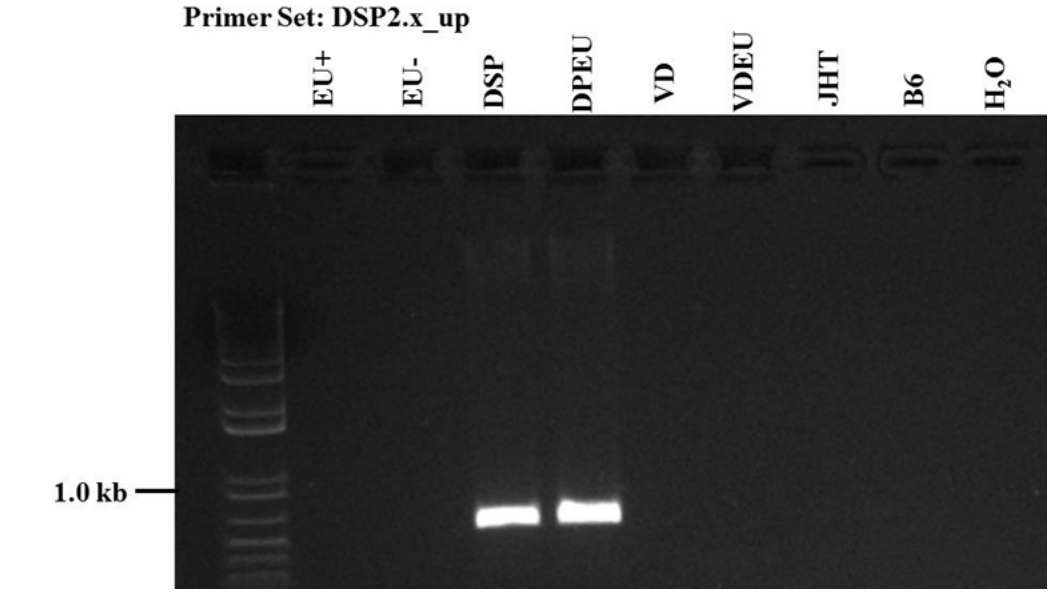

Primer Set: Eμ 5F, Eμ 3'
Primer Set: Eμ 5F, Eμ 2R
EU+
EU-
DSP
DPEU
VD
VDEU
JHT
B6
1.0 kb
500 bp

Figure 9

C


Primer Set: DSP2.x_up
EU+
EU-
DSP
DPEU
VD
VDEU
JHT
B6
H2O
1.0 kb

D

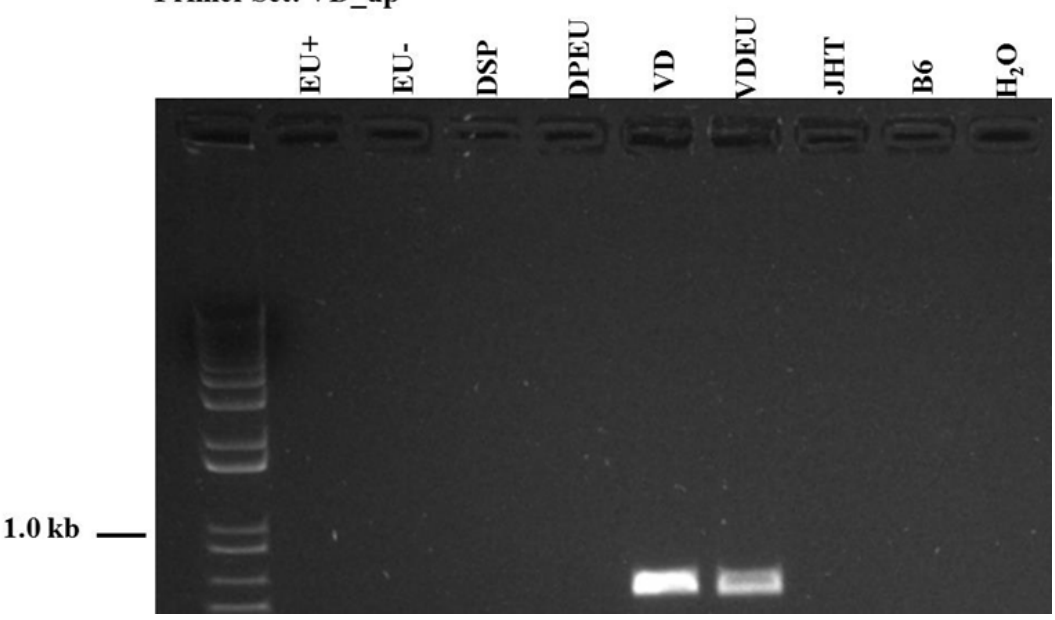
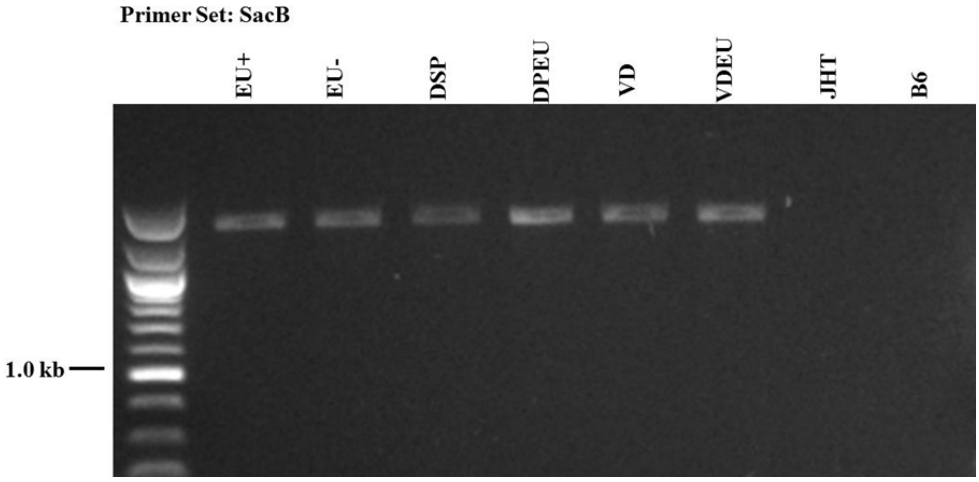

Primer Set: VD_up
EU+
EU-
DSP
DPEU
VD
VDEU
JHT
B6
H2O
1.0 kb

Figure 9

E


Primer Set: SacB
EU+
EU-
DSP
DPEU
VD
VDEU
JHT
B6
1.0 kb

F

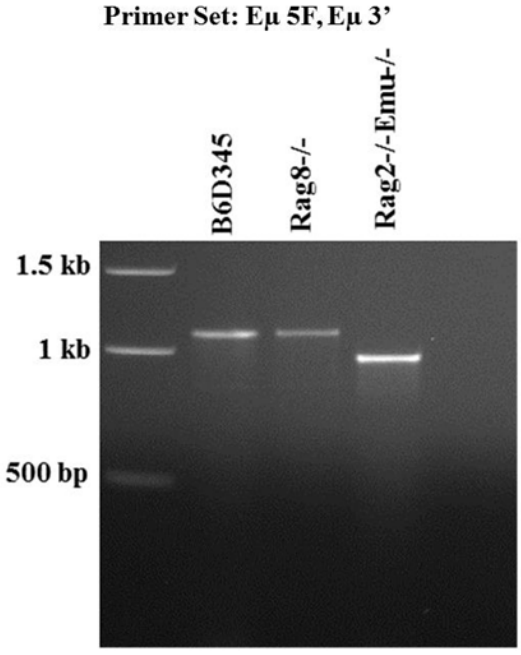

Primer Set: Eμ 5F, Eμ 3'
B6D345
Rag8-/-
Rag2-/-Emu-/-
1.5 kb
1 kb
500 bp

Figure 9. Genotyping Overview to Test Presence of Transgene in BAC Constructs. The overall genotyping strategy is given in the illustration and 1% agarose gels above. The BAC constructs are derived and modified from a 181 kb sequence on mouse chromosome 12 (chr12:114,567,000 – 114,748,000) (Figure 11). In (A), the basic PCR strategy is illustrated as it relates to the Eμ enhancer on mouse chromosome 12. The "outside" genotyping primers, "Eμ 5F" and "Eμ 3,'" yields a 1.1 kb PCR band (chr12: 114,665,284 - 114,666,383) when the Eμ enhancer is in its normal location (EU+, DPEU, VDEU). The "inside" genotyping primers, "Eμ 5F" and "Eμ 2R,'" detect whether Eμ is present or absent (EU-, DSP, VD), since the Eμ 2R is a reverse primer that is upstream of the Hinf I enzyme that defines the 3' end of the Eμ core region. If the Eμ enhancer is present as tested with the "inside" primer set, then a 650 bp product results (EU+, DPEU, VDEU) (chr12: 114,665,678 – 114,666,329). Both sets of primers are depicted in (B), with the "outside" set represented on the left gel, and the "inside" set represented on the right gel." To detect the relocated Eμ enhancer in the DSP and DPEU transgene, a forward primer (chr12: 114,697,256-114,697,275) that is about 2 kb from DSP2.X (chr.12:114,699,580 – 114,699,720), and a reverse primer inside the Eμ enhancer yields a 800 bp product (C). This primer pair is referred to as "DSP2.x_up." To test for the presence of the Eμ enhancer in the VD and VDEU transgenes, a forward primer directly upstream of the relocated Eμ enhancer (chr12: 114,725,356 - 114,725,375), and a reverse primer inside the Eμ enhancer yield a 750 bp product (D). This primer pair is referred to as "VD_up." To determine the presence of the transgene independent of the Eμ enhancer, a specific primer set that corresponds to the SacB gene, which is part of the vector, yields a 1.2 kb product (E). For reference, a 1% agarose gel

of the cell lines is shown in (F), where the Eμ enhancer is present in $Rag8^{-/-}$ and B6D345 to yield a 1045 bp product, but the 220 bp Eμ-core that is located between the two Hinf I sites (chr12:114665614 – 114665833) is deleted in $Rag2^{-/-}$ $E\mu^{-/-}$.

Figure 10

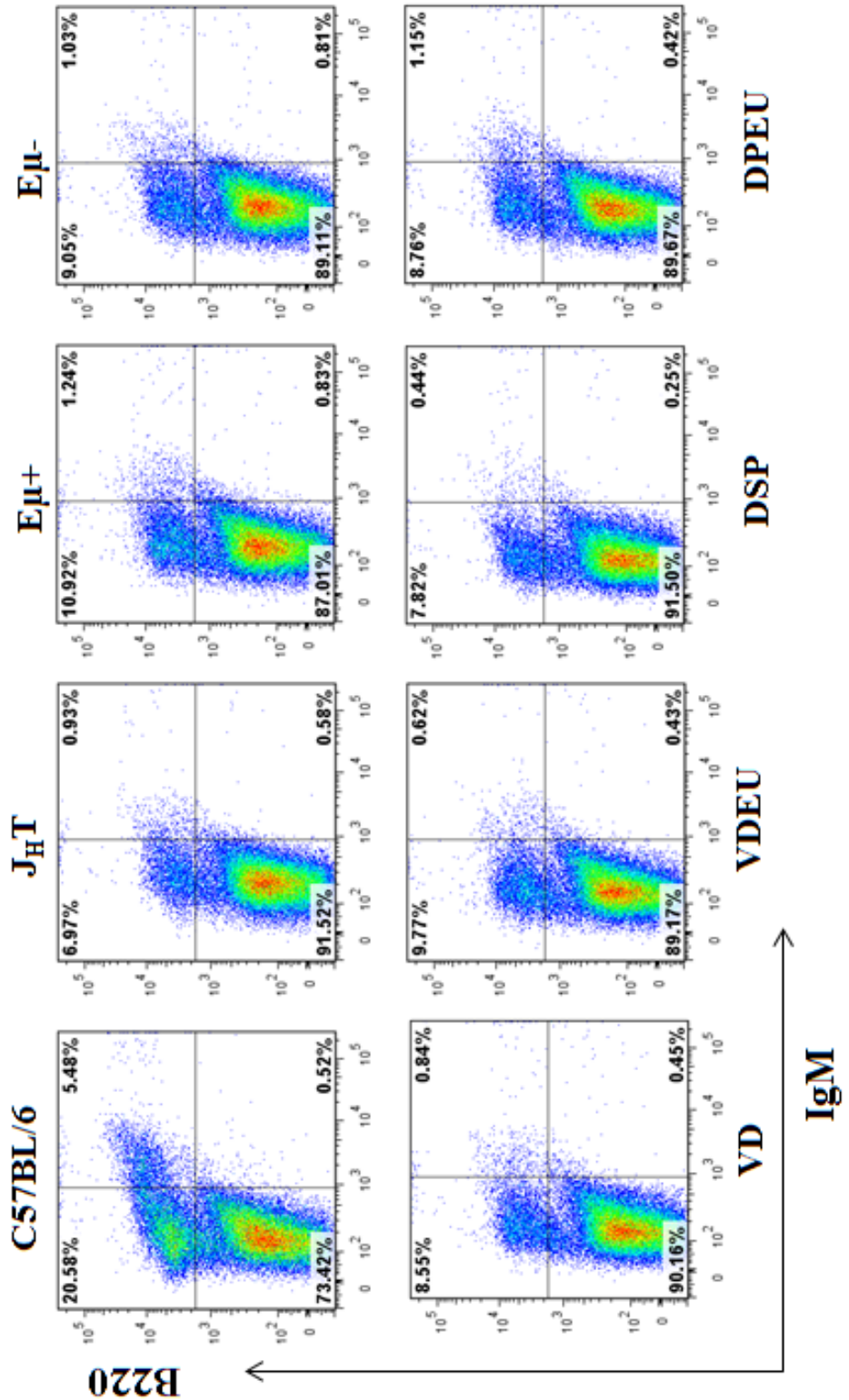

C57BL/6
JHT
Eμ+
Eμ-
VD
VDEU
DSP
DPEU
B220
IgM
20.58%
5.48%
73.42%
0.52%
6.97%
0.93%
91.52%
0.58%
10.92%
1.24%
87.01%
0.83%
9.05%
1.03%
89.11%
0.81%
8.55%
0.84%
90.16%
0.45%
9.77%
0.62%
89.17%
0.43%
7.82%
0.44%
91.50%
0.25%
8.76%
1.15%
89.67%
0.42%

Figure 10. FACs analysis to confirm J_HT background. Measurement of $B220^+$ and IgM^+ bone marrow cells. Single-cell suspensions of bone marrow cells from femora and tibiae of transgenic mice were resuspended in ACK lysis buffer (155 mM NH_4Cl, 10 mM $KHCO_3$, 100 μM EDTA) for 1 min at room temperature to lyse red blood cells before staining in FACS buffer (PBS, 0.5% BSA, 0.01% NaN3, pH 7.2-7.4) containing the appropriate, pre-titered B-cell marker antibodies (B220 FITC, BD catalog #, 561876; IgM APC, BD catalog # 551062).

Figure 11

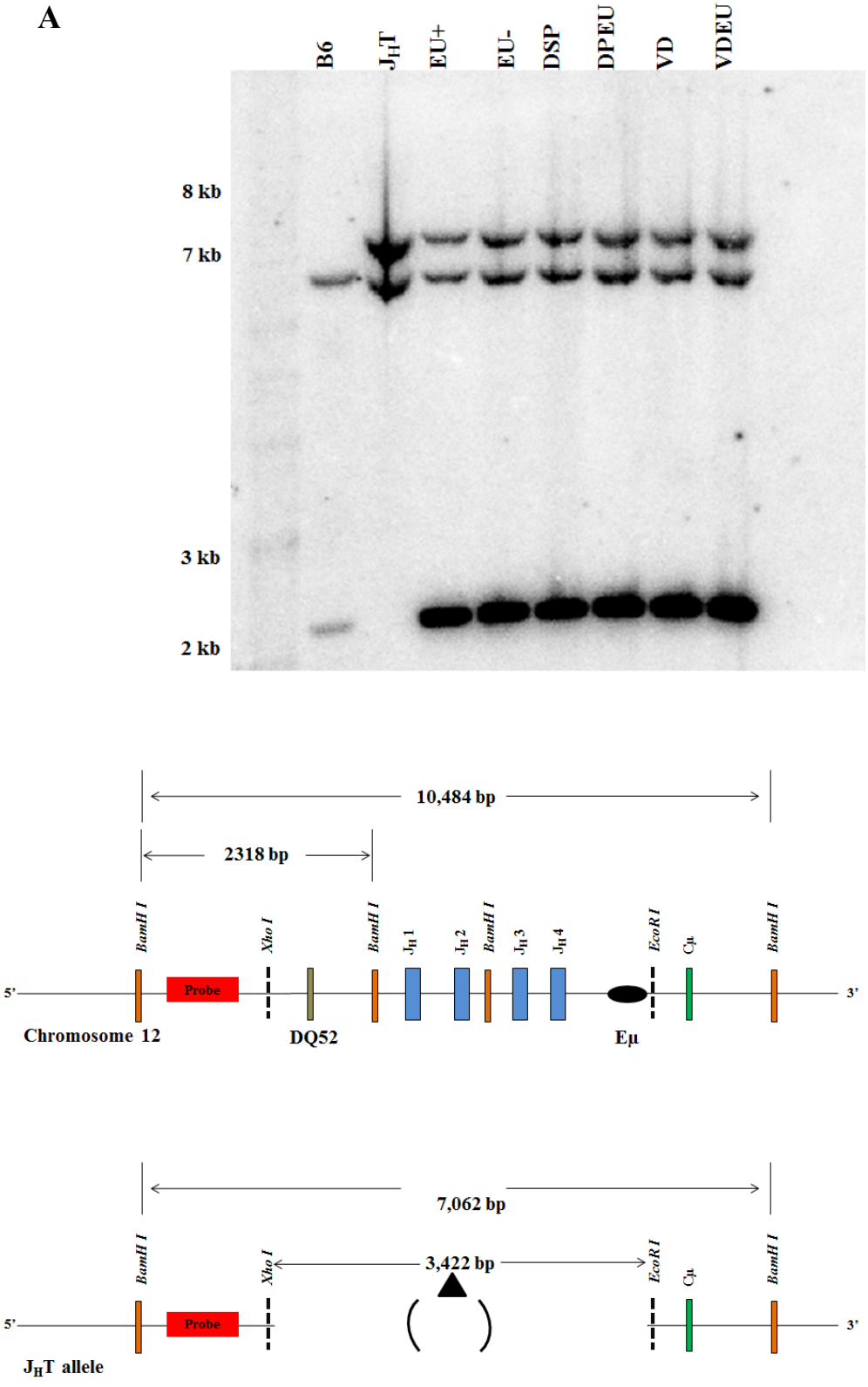

A
B6
JHT
EU+
EU-
DSP
DPEU
VD
VDEU
8 kb
7 kb
3 kb
2 kb
10,484 bp
2318 bp
BamH I
Xho I
BamH I
JH1
JH2
BamH I
JH3
JH4
EcoR I
Cμ
BamH I
5'
Probe
3'
Chromosome 12
DQ52
Eμ
7,062 bp
3,422 bp
BamH I
Xho I
EcoR I
Cμ
BamH I
5'
Probe
3'
JHT allele

Figure 11

B

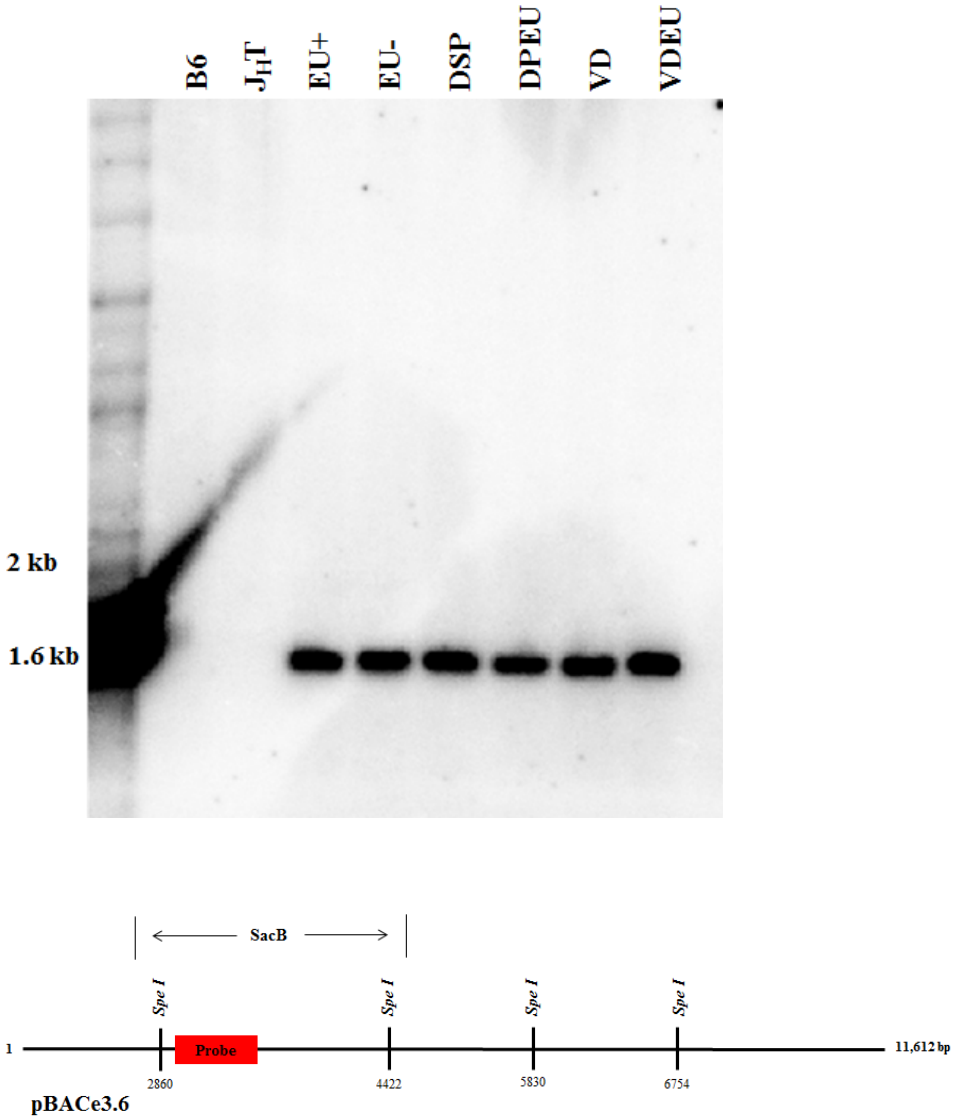

B6
J_HT
EU+
EU-
DSP
DPEU
VD
VDEU
2 kb
1.6 kb
SacB
Spe I
Spe I
Spe I
Spe I
1
Probe
11,612 bp
2860
4422
5830
6754
pBACe3.6

Figure 11. Southern Blot to Confirm Presence of Transgene and Relative Copy Number. Southern blots were performed to confirm the presence of the transgene on the JHT background, and relative copy number. The southern blot in (A) shows the transgene in the bottom band and JHT background in the top band. Probe used to detect transgene on JHT background is in red color in the schematic and is described in the Materials and Methods and is from chr12: 114,669,611 – 114,670,131. Kidney DNA samples were digested with BamH I, which is shown in orange bars in the schematic. This enzyme has four cut sites on wild-type alleles: one 2 kb upstream of DQ52 and outside of the JHT deletion, one between DQ52 and JH1, one between JH1 and JH2, and one 5 kb downstream of the Eμ enhancer. Transgene samples on a JHT background yield two bands, one at 7,062 bp that corresponds to the JHT background, and one at 2,318 that marks the BAC transgene as it reflects mouse chromosome 12: 114,668,479 – 114,670,802. The southern blot in (B) is for a probe that detects only the SacB sequence when the transgene is cut with Spe I enzyme to yield a 1,562 bp band that is on the backbone of the BAC, which illustrated beneath the gel. In both gels, the relative copy number was estimated to be 4-5 fold different from the endogenous copy.

Figure 12

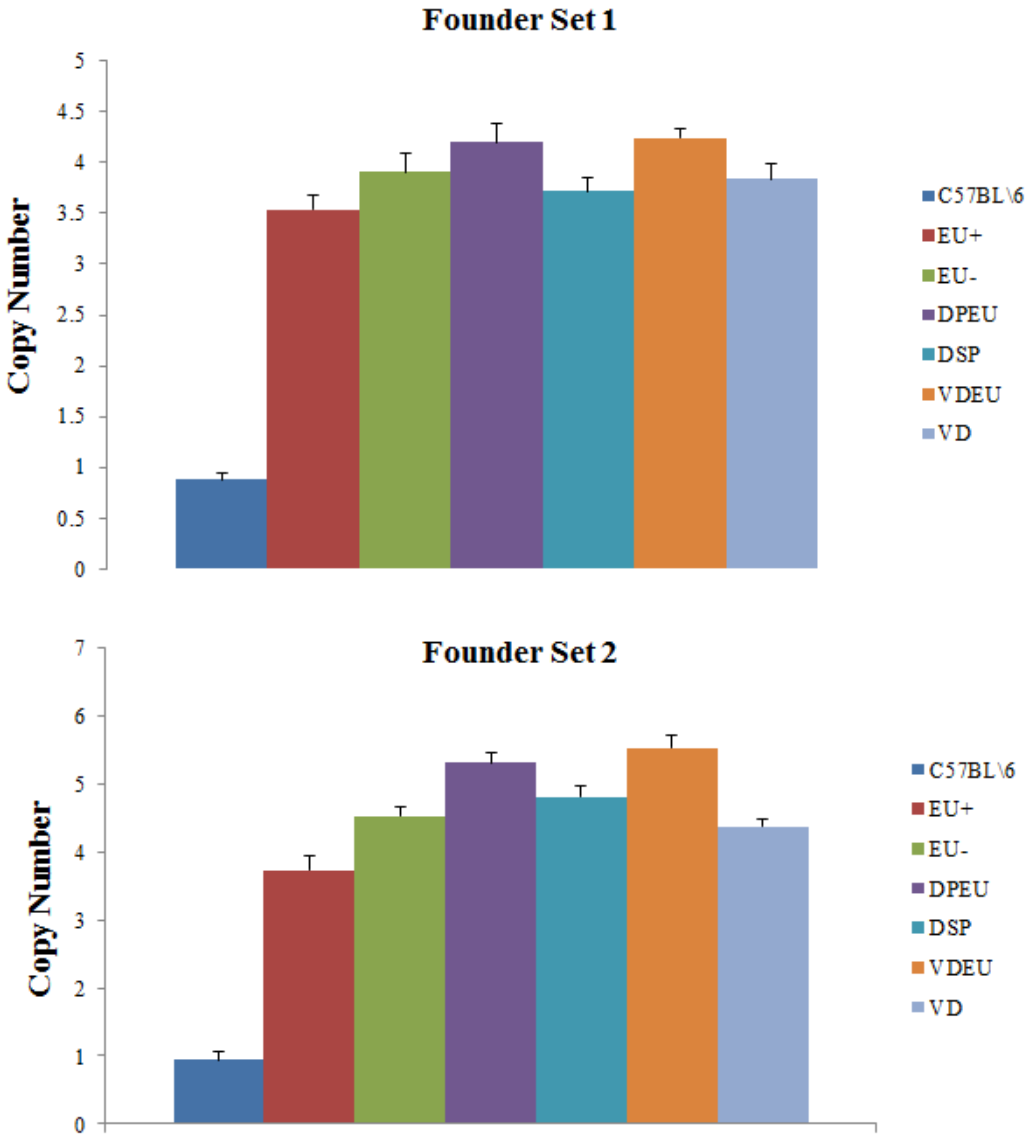
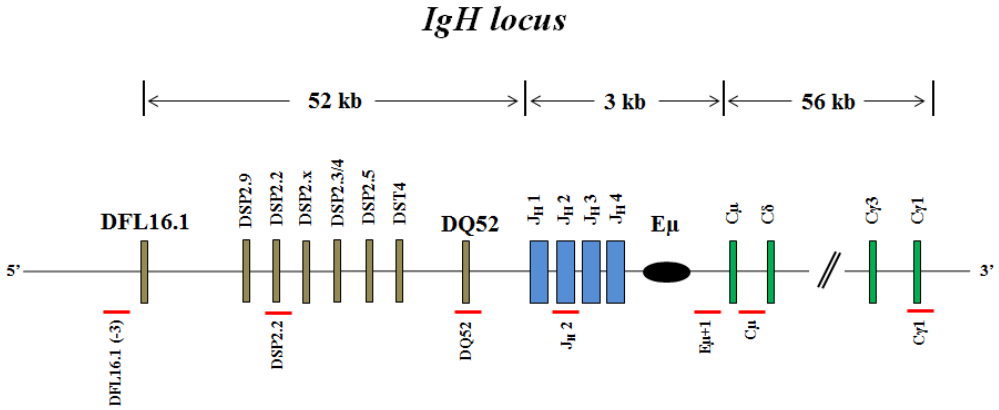

Founder Set 1
Copy Number
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
C57BL\6
EU+
EU-
DPEU
DSP
VDEU
VD
Founder Set 2
Copy Number
7
6
5
4
3
2
1
0
C57BL\6
EU+
EU-
DPEU
DSP
VDEU
VD

Figure 12. Copy Number in Transgenic Mice. Mouse-specific qPCR copy number assays (qBiomarker Copy Number PCR Assays) were purchased from Qiagen. A mouse multicopy reference assay (Qiagen, Catalog #. VPM000-0000000A) was used to normalize input DNA. Primers were specific for mouse chromosome 12:114,630,601 – 114,630,800 (Qiagen, Catalog #. VPM112-0573154A). Copy number was similar within each founder set. For the EU+ strain, only a single founder was used for all experiments.

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

A

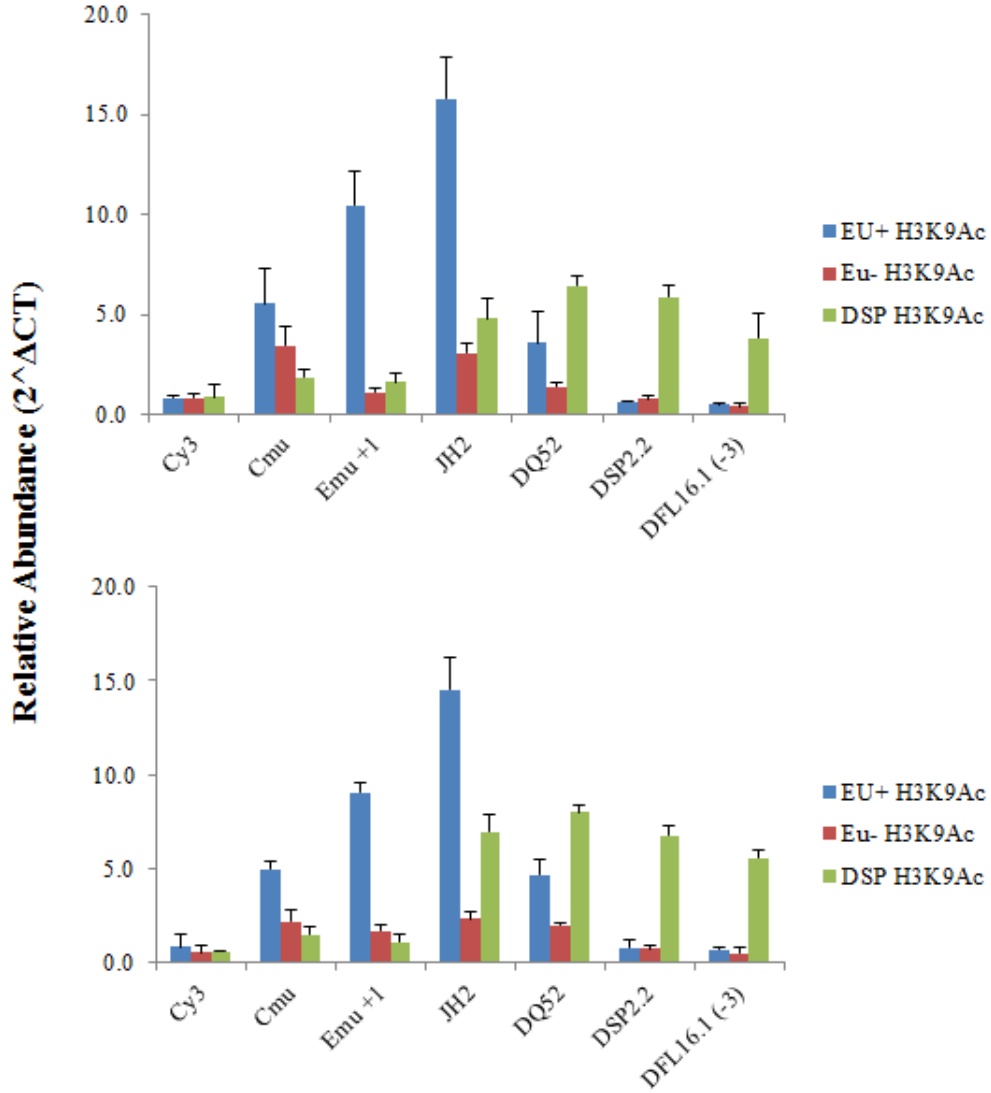
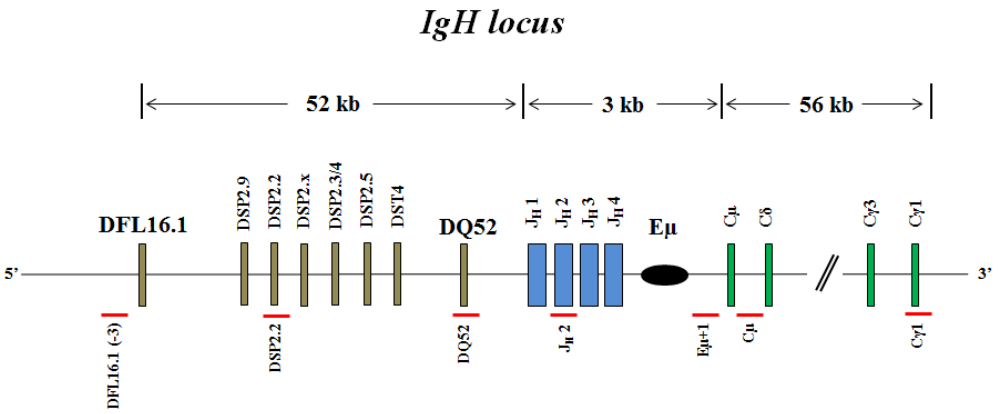

Relative Abundance (2^ΔCT)
20.0
15.0
10.0
5.0
0.0
Cy3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K9Ac
Eu- H3K9Ac
DSP H3K9Ac

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

B

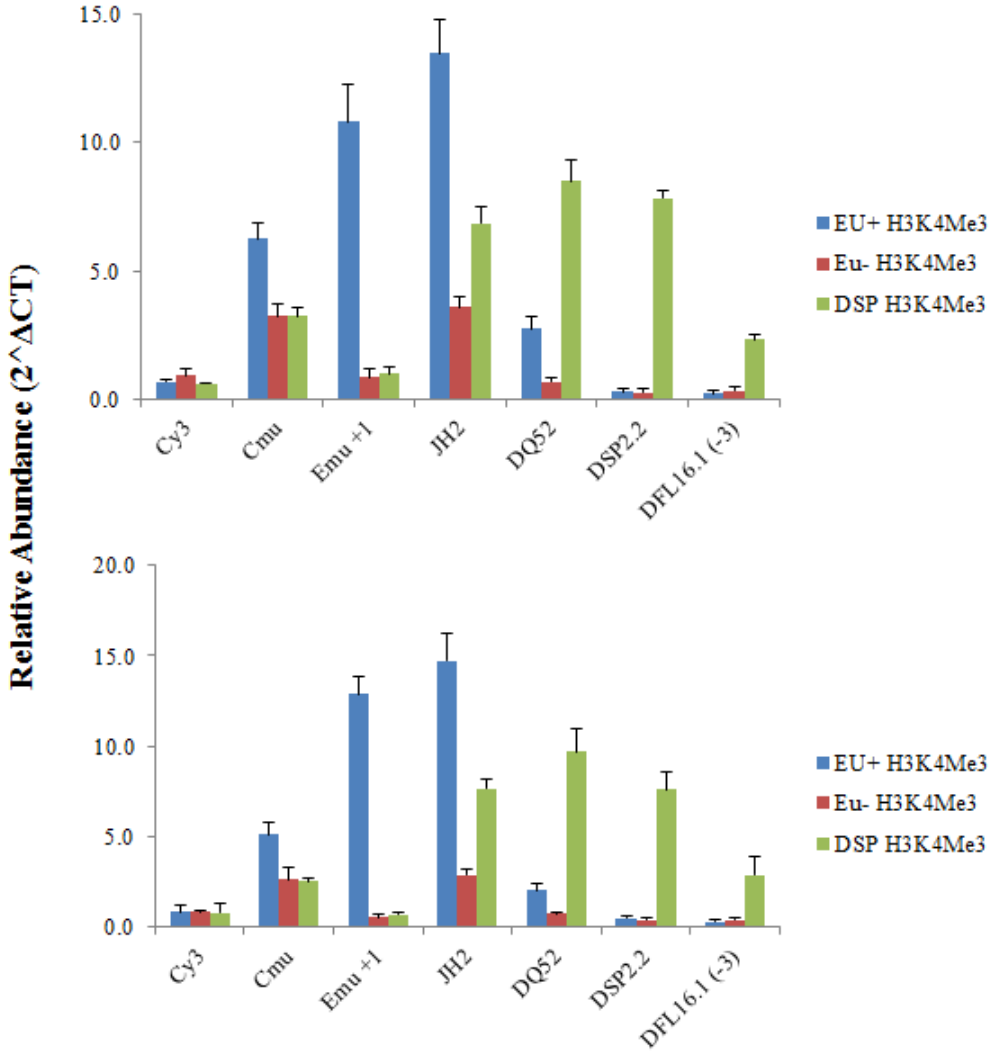
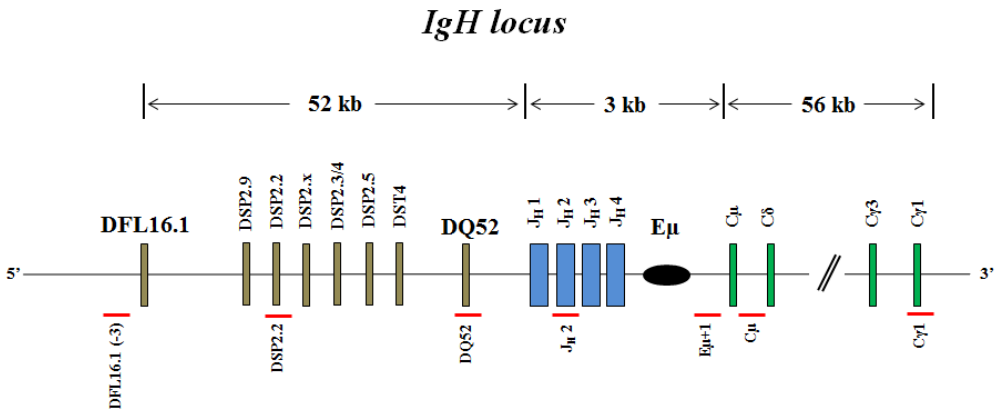

Relative Abundance (2^ΔCT)
15.0
10.0
5.0
0.0
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K4Me3
Eu- H3K4Me3
DSP H3K4Me3
20.0
15.0
10.0
5.0
0.0
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K4Me3
Eu- H3K4Me3
DSP H3K4Me3

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

C

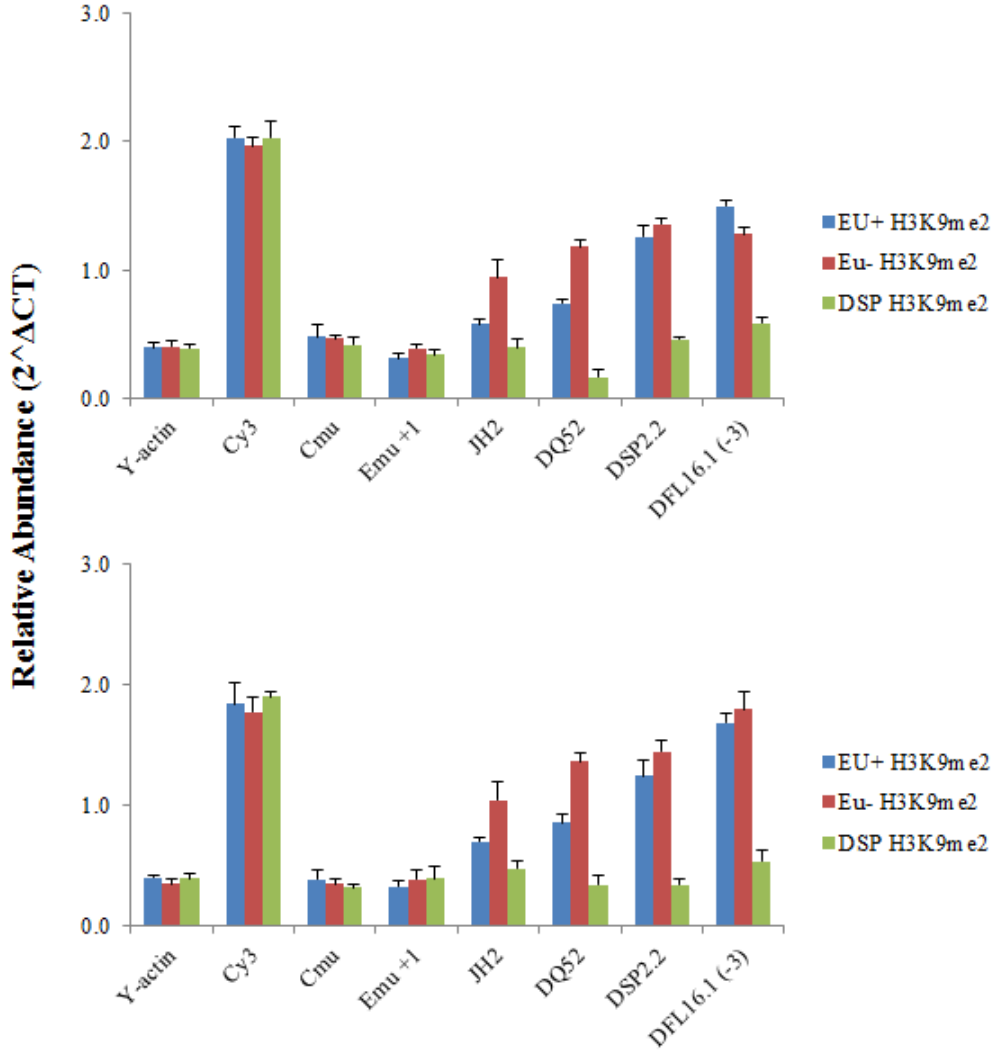
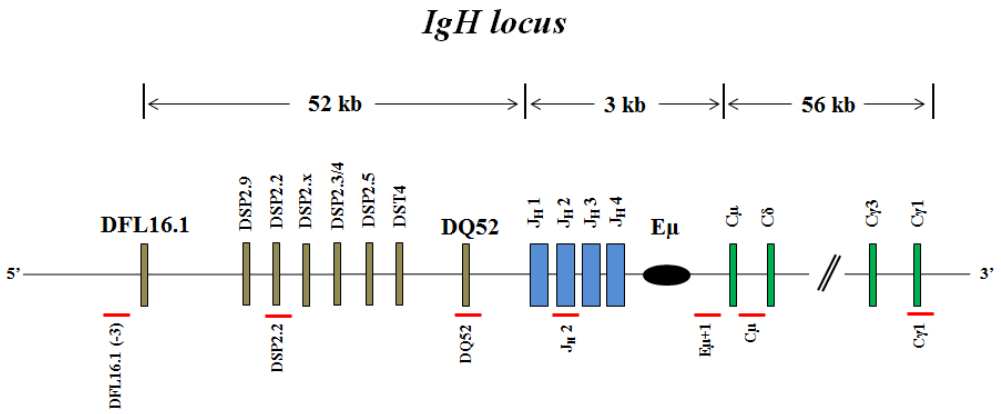

Relative Abundance (2^ΔCT)
3.0
2.0
1.0
0.0
EU+ H3K9m e2
Eu- H3K9m e2
DSP H3K9m e2
Y-actin
Cy3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

D

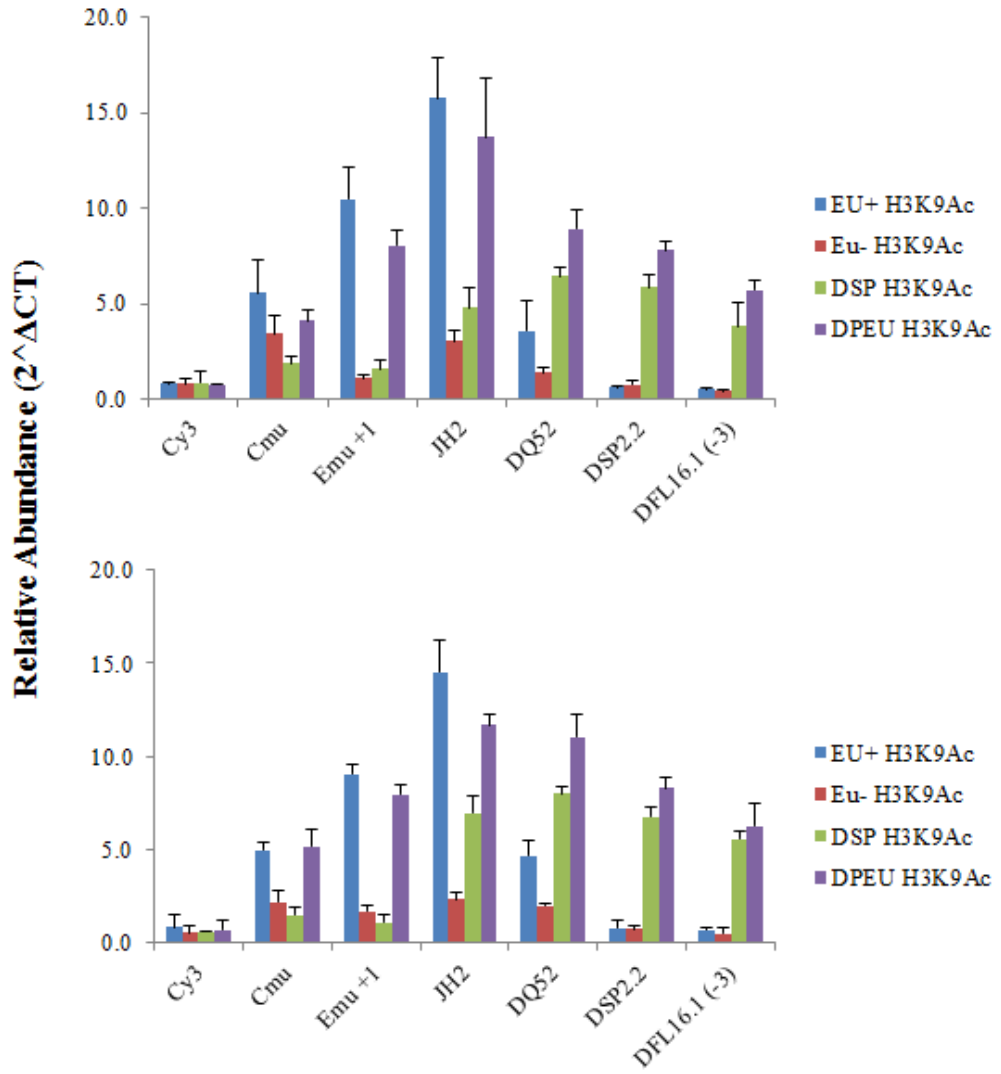
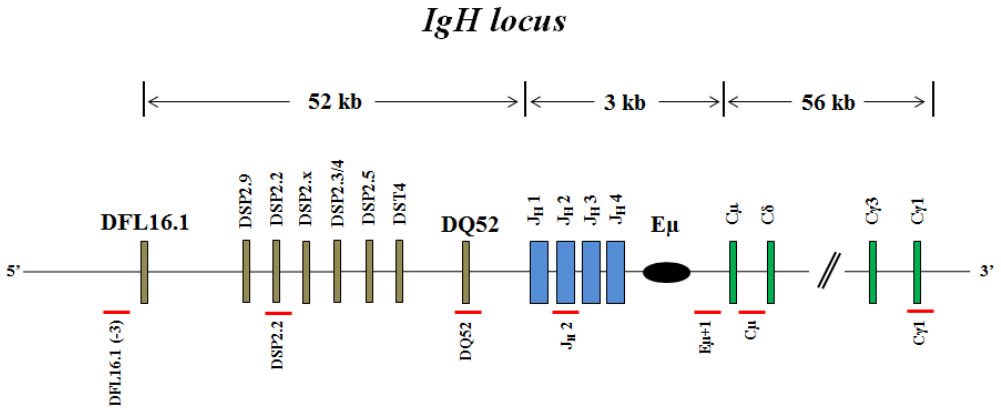

Relative Abundance (2^ΔCT)
20.0
15.0
10.0
5.0
0.0
Cy3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K9Ac
Eu- H3K9Ac
DSP H3K9Ac
DPEU H3K9Ac

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

E

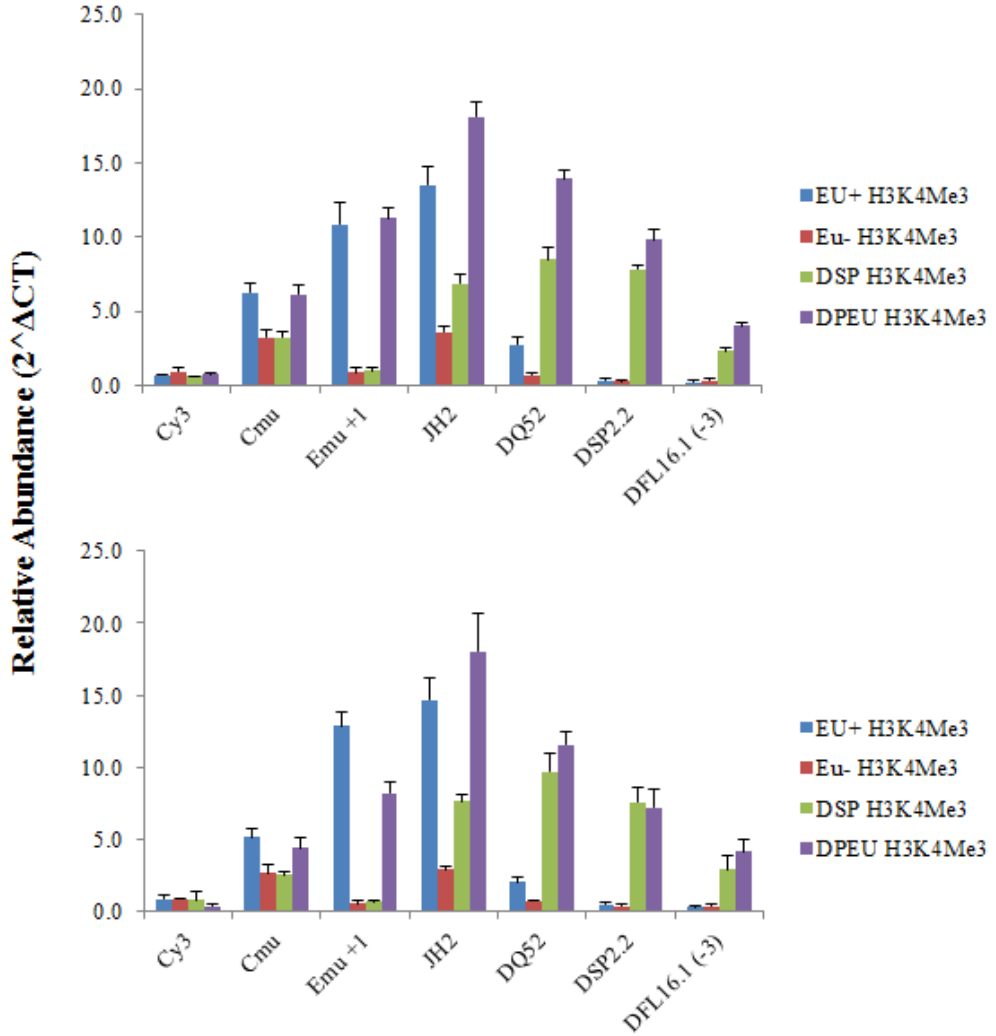
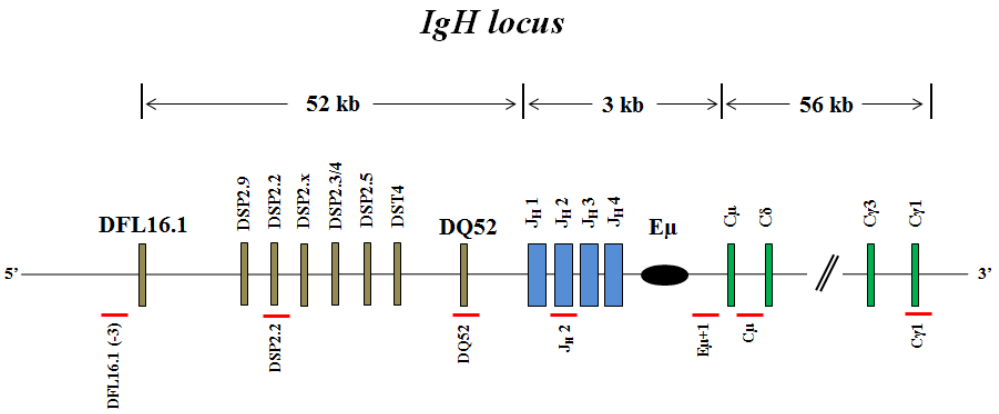

Relative Abundance (2^ΔCT)
25.0
20.0
15.0
10.0
5.0
0.0
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K4Me3
Eu- H3K4Me3
DSP H3K4Me3
DPEU H3K4Me3

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

F

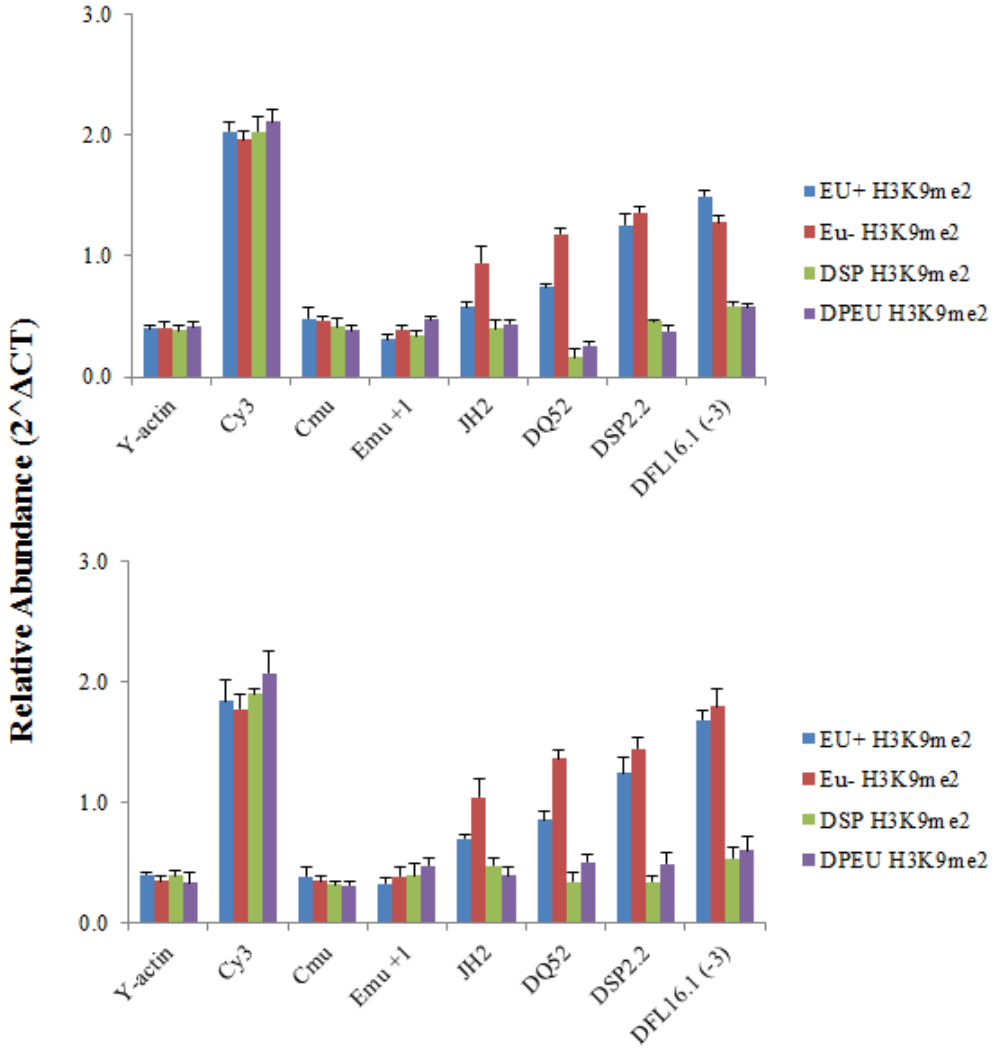
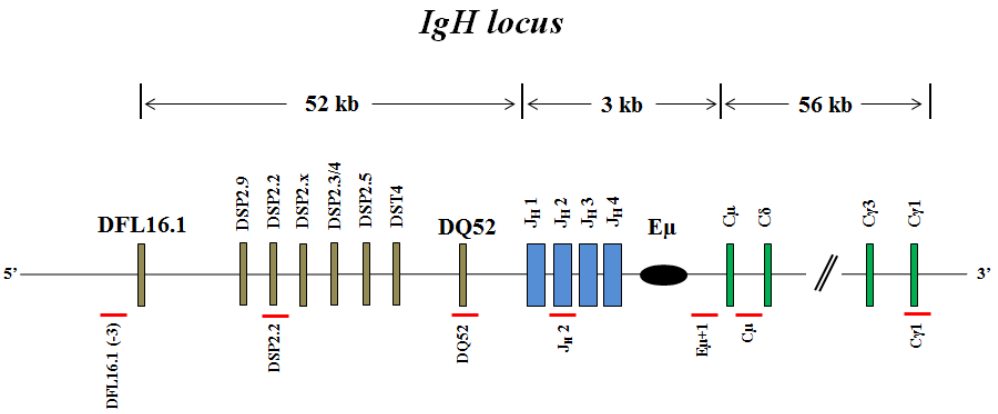

Relative Abundance (2^ΔCT)
3.0
2.0
1.0
0.0
EU+ H3K9me2
Eu- H3K9me2
DSP H3K9me2
DPEU H3K9me2
Y-actin
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

G

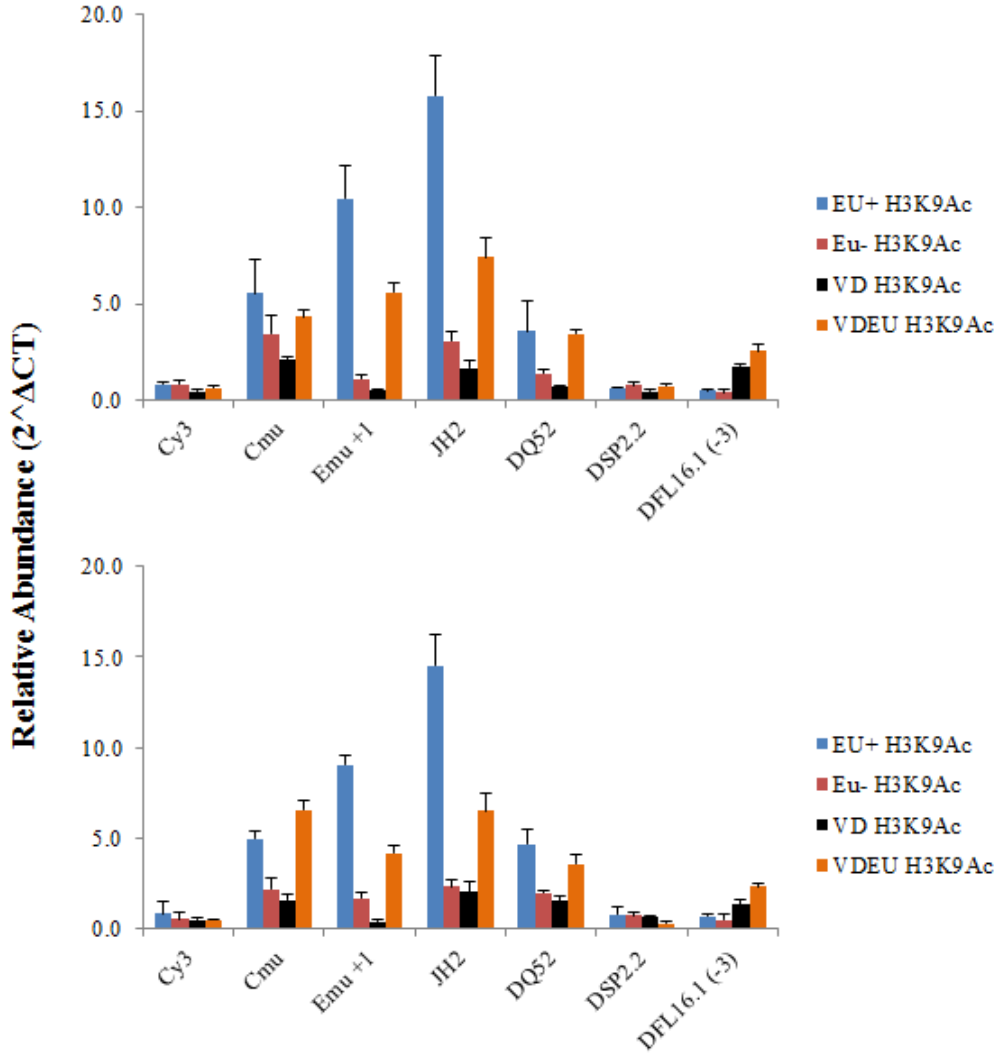
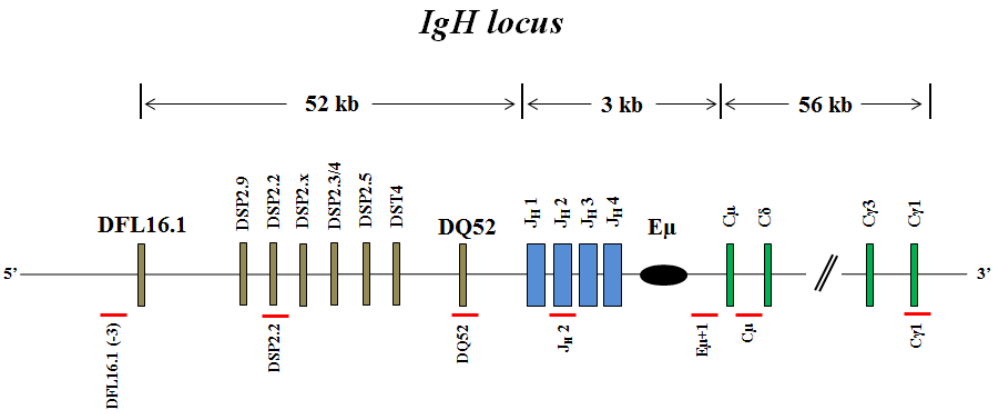

Relative Abundance (2^ΔCT)
20.0
15.0
10.0
5.0
0.0
EU+ H3K9Ac
Eu- H3K9Ac
VD H3K9Ac
VDEU H3K9Ac
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

H

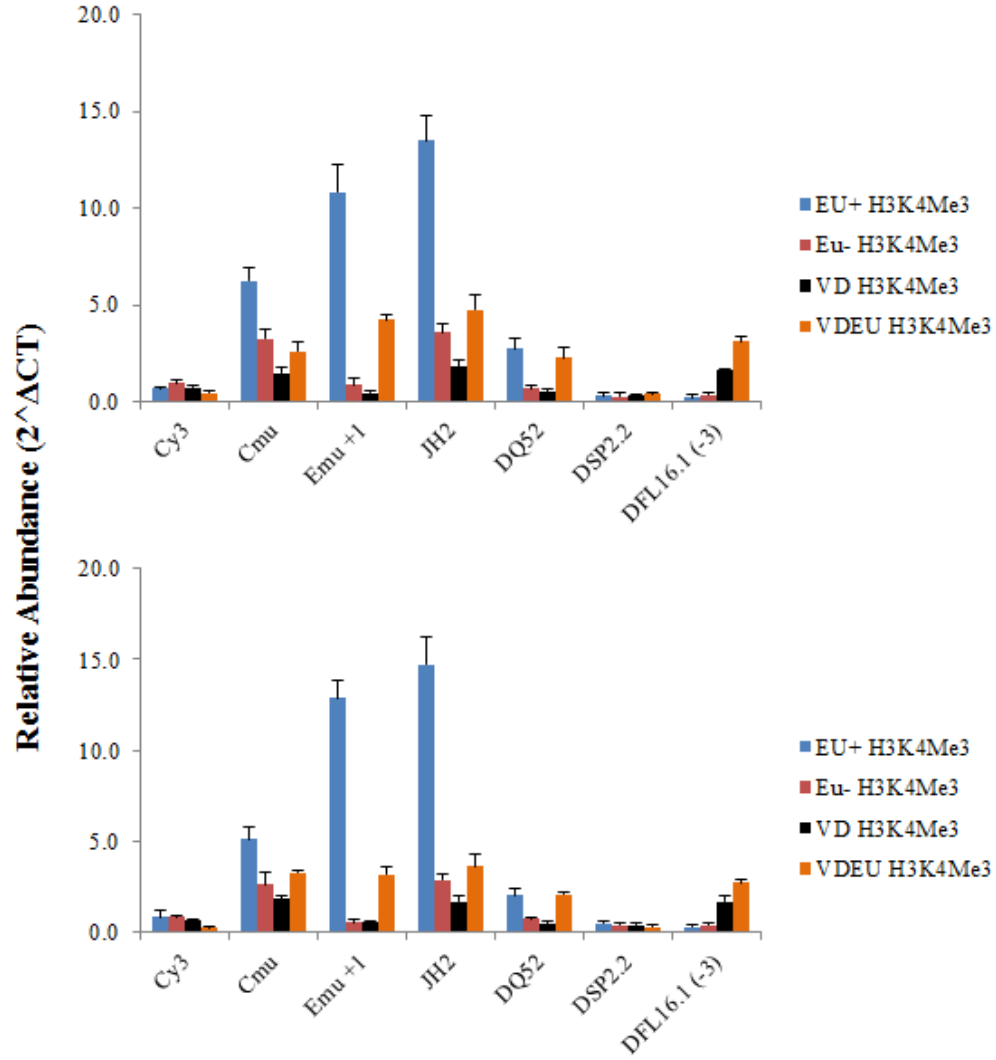
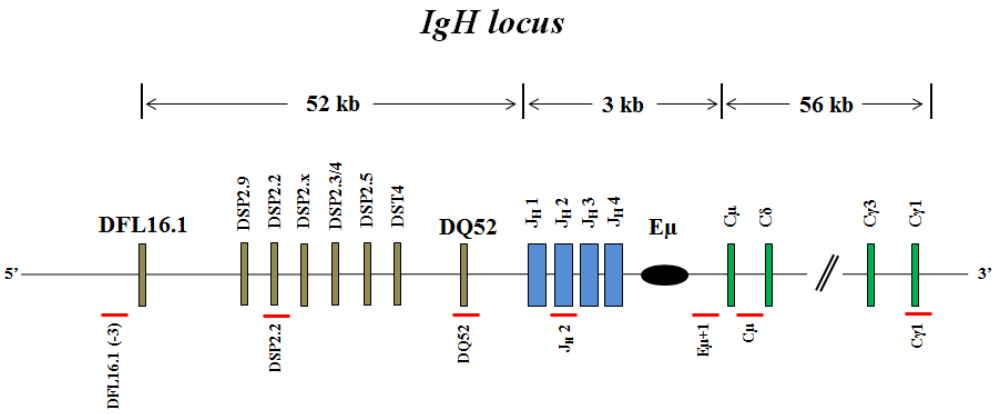

Relative Abundance (2^ΔCT)
20.0
15.0
10.0
5.0
0.0
Cγ3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)
EU+ H3K4Me3
Eu- H3K4Me3
VD H3K4Me3
VDEU H3K4Me3

Figure 13


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH 1
JH 2
JH 3
JH 4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-3)
DSP2.2
DQ52
JH 2
Eμ+1
Cμ
Cγ1

I

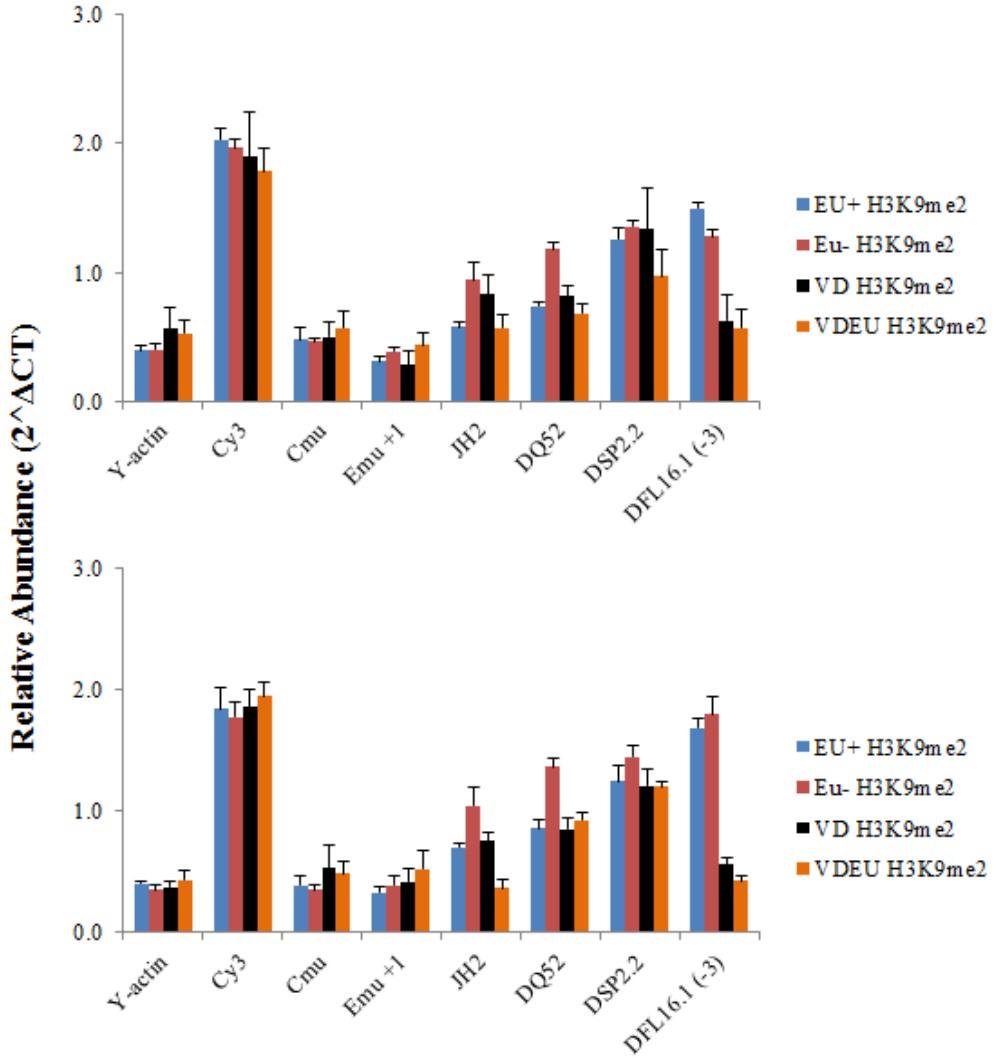

Relative Abundance (2^ΔCT)
3.0
2.0
1.0
0.0
EU+ H3K9me2
Eu- H3K9me2
VD H3K9me2
VDEU H3K9me2
Y-actin
Cy3
Cmu
Emu +1
JH2
DQ52
DSP2.2
DFL16.1 (-3)

Figure 13. Eμ Dependent Histone Modifications in BAC Transgenes. CD19+ bone marrow pro-B cells of BAC transgenic mice were grown on OP-9 cultures for 9 - 11 days and used in chromatin immunoprecipitation (ChIP) assays using anti-H3K9ac (A, D, G), anti-H3K4me3 (B, E, H), or anti-H3K9me2 antibodies (C, F, I). After formaldehyde crosslinking the cells, the crosslinked protein-DNA complex was extracted by lysis to produce nuclear material that was subsequently sheared by sonication to yield only protein binding sequences. Referenced antibodies were then used to selectively enrich for the protein-DNA complex of interest. Crosslinks were then reversed and the DNA was purified and quantified fluorometrically for qPCR. A total of 200 pg was used in each qPCR reaction. Experiments pooled six mice of each genotype of separate founders. Similar results were obtained for each founder set (Materials and Methods). Data from Founder Set 1 is displayed on top, and data from Founder Set 2 is displayed on bottom. Positions of amplicons are displayed by the red bars in the IgH schematic at the top of each figure. Results shown are from two independent cell preparations and immunoprecipitates analyzed in duplicates. Error bars represent standard deviation between experiments.

Figure 14

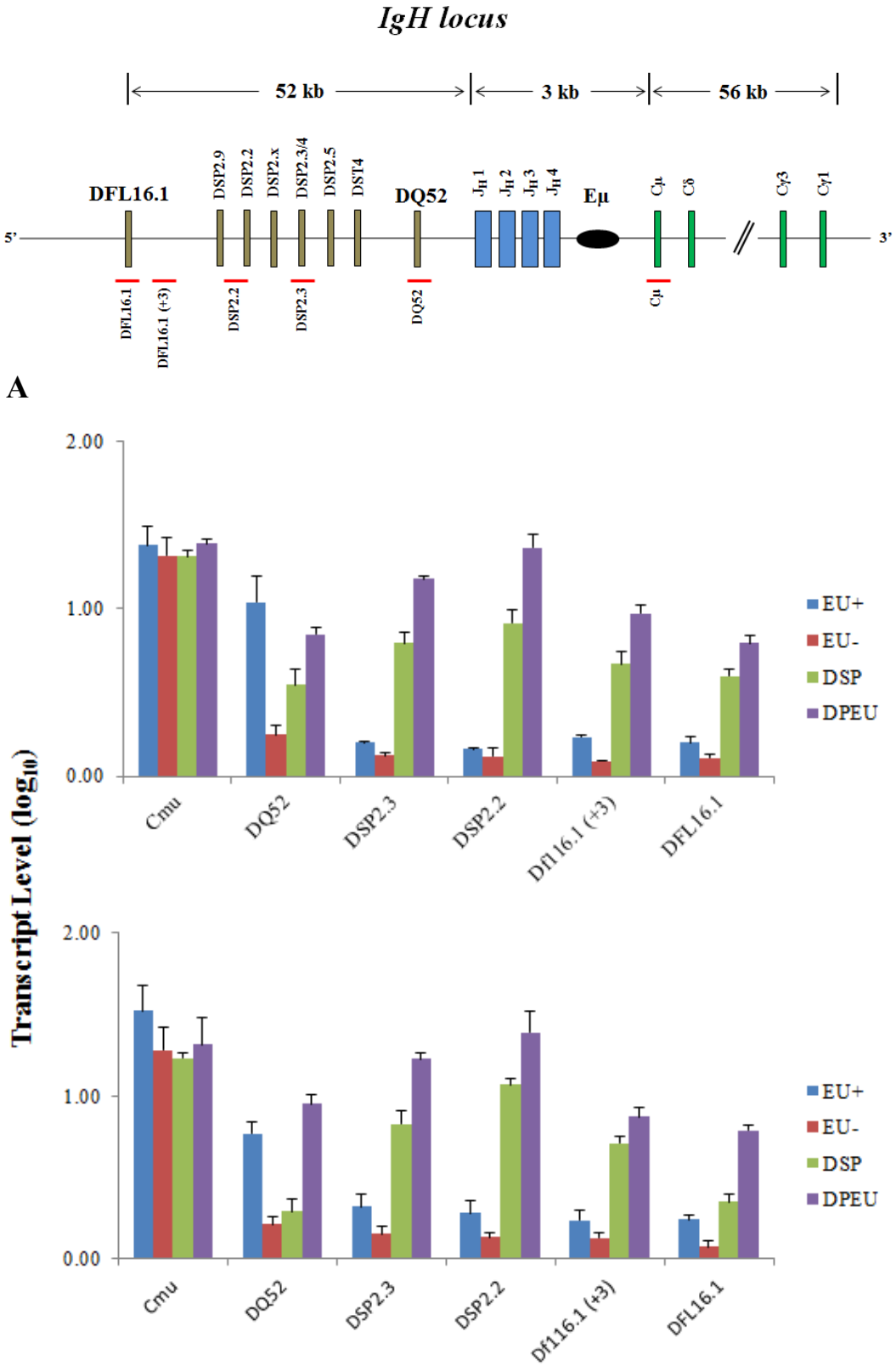
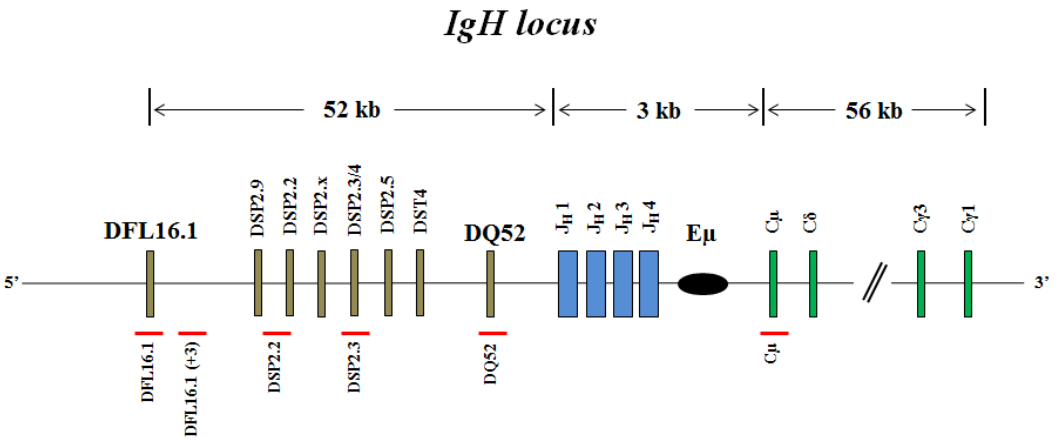

IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1
DFL16.1 (+3)
DSP2.2
DSP2.3
DQ52
Cμ
A
Transcript Level (log10)
2.00
1.00
0.00
EU+
EU-
DSP
DPEU
Cmu
DQ52
DSP2.3
DSP2.2
Df116.1 (+3)
DFL16.1
Df116.1 (+3)

Figure 14


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1
DFL16.1 (+3)
DSP2.2
DSP2.3
DQ52
Cμ

B

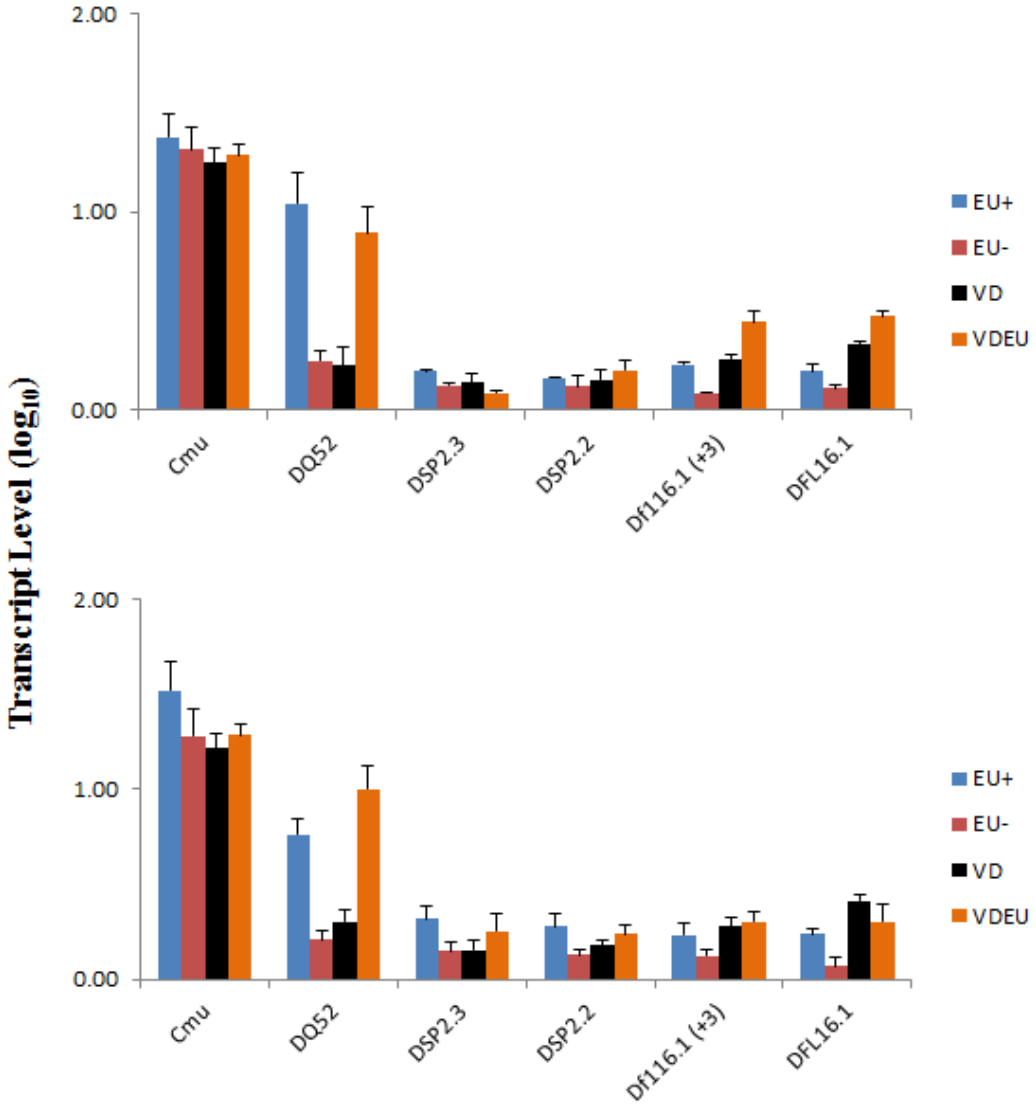
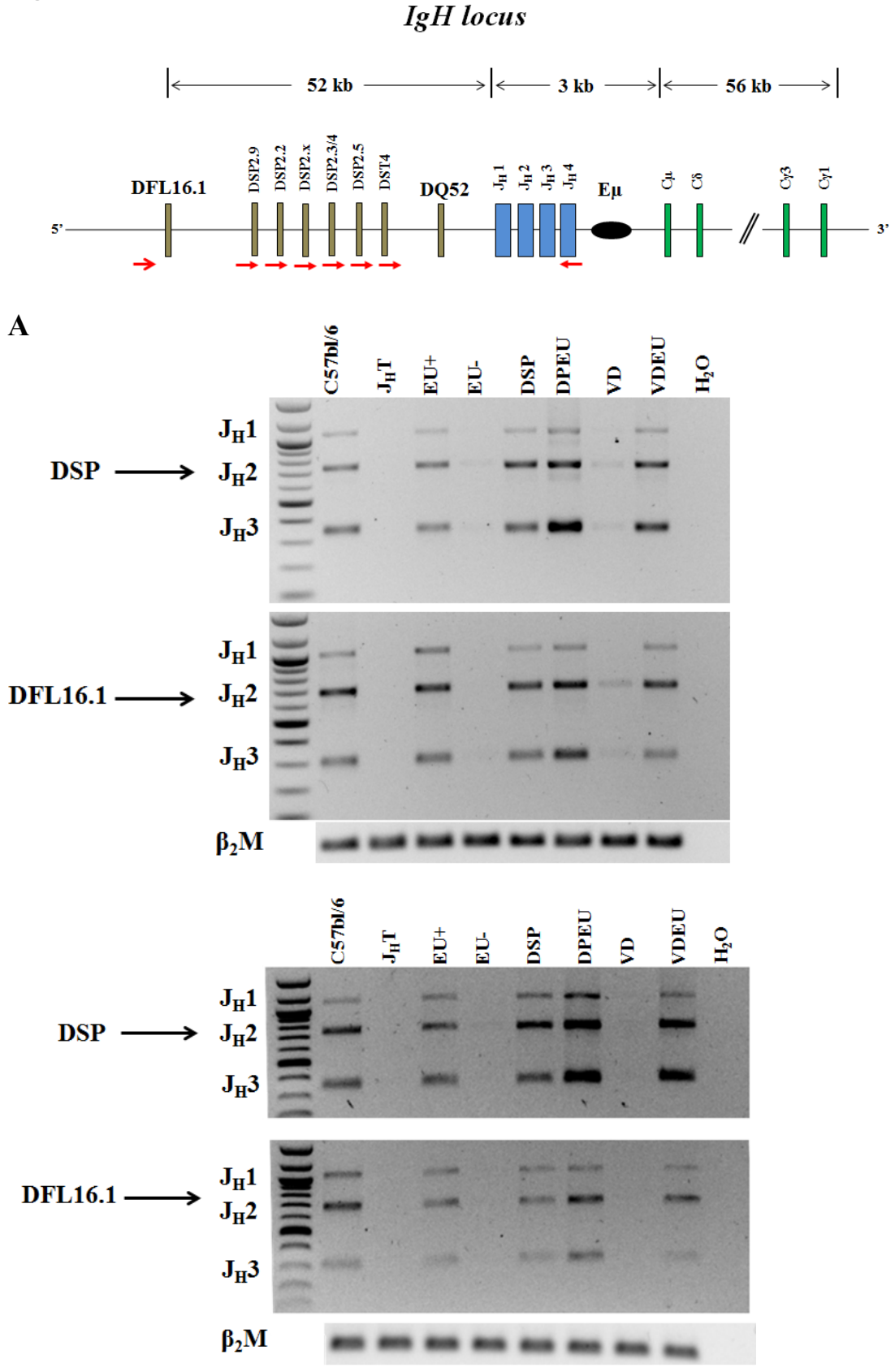

Transcript Level (log10)
2.00
1.00
0.00
EU+
EU-
VD
VDEU
Cmu
DQ52
DSP2.3
DSP2.2
Df116.1 (+3)
DFL16.1

Figure 14. Eμ Dependent Histone Transcription in BAC Transgenes. Total RNA was obtained from primary bone-marrow pro-B cells of each transgenic mouse strain and converted into complementary DNA using random hexamers and reverse transcription (Invitrogen, catalog # 10928-034). Samples were quantitated and assayed by qPCR for amplicons that are shown in the schematic at the top of this figure. To compare between transgenes, the data with each primer pair was normalized to γ-actin expression level. Data shown is the mean of two independent RNA isolations from each transgenic strain in duplicates with each primer set for each founder. Figure (A) shows transcripts for EU +, EU-, DSP, and DPEU. Figure (B) shows transcripts for VD and VDEU. Founder 1 is at the top, and Founder 2 is at the bottom of each figure. Error bars indicate the standard deviation between experiments.

Figure 15


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
A
C57bl/6
JHT
EU+
EU-
DSP
DPEU
VD
VDEU
H2O
DSP
DFL16.1
JH1
JH2
JH3
β2M

B

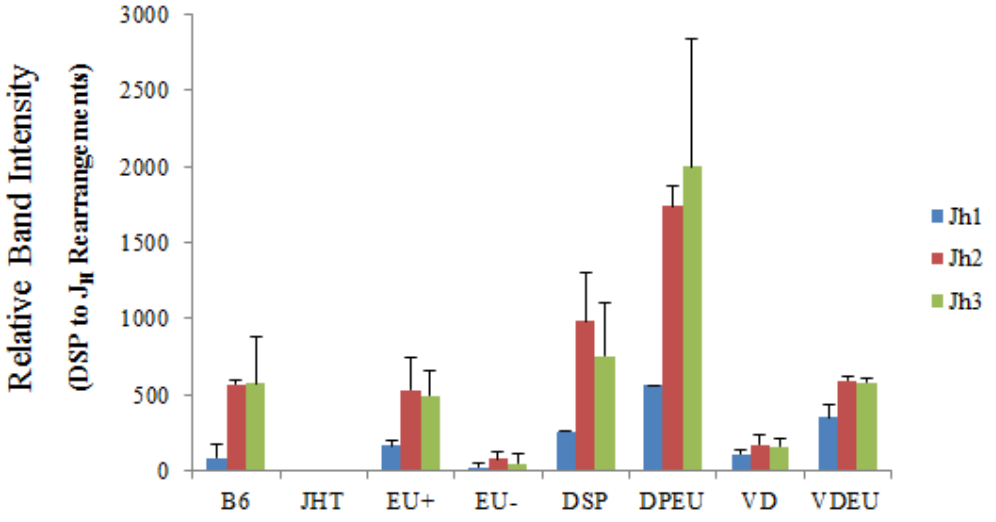
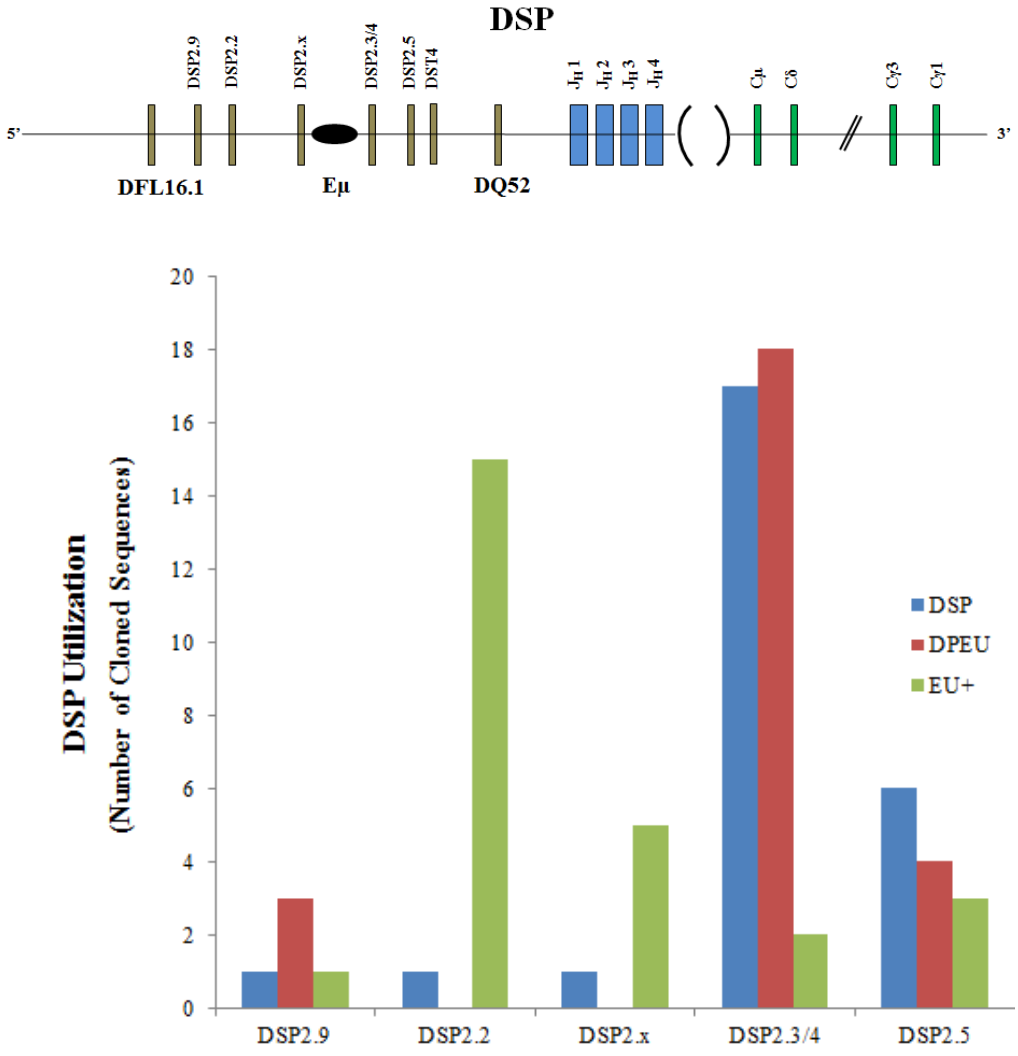

Relative Band Intensity
(DSP to J_H Rearrangements)
3000
2500
2000
1500
1000
500
0
B6
JHT
EU+
EU-
DSP
DPEU
VD
VDEU
Jh1
Jh2
Jh3

Figure 15. Eμ Induced D_H to J_H Gene Recombination in Transgenic Mice. (A) PCR analysis of D_H to J_H gene rearrangements in primary $CD19^+$ pro-B cells from the bone marrow of each transgenic mouse. β2-microglobulin used as a loading control to ensure equal loading of DNA from all samples. Genomic DNA was prepared from 6 mice of each strain and used to analyze DSP2 and DFL16.1 rearrangements as described in the Material and Methods section. Data from founder set 1 is on top and founder set 2 is on the bottom. Location of the D_H specific forward primers and the common J_H reverse primer are shown in red bars under the schematic at the top of the figure. Following PCR, the products were fractionated through a 1% agarose gel for 45 minutes, 120V and analyzed on Syngene's G:BOX F3 Gel Imager. The relative band intensity was measured for the DSP to J_H gene rearrangements for founder set 1 of each transgenic strain in (B), quantified by using the Gene Tools Syngene program. Data shown is representative of two independent experiments and error bars represent the standard deviation between experiments.

Figure 16

A


DSP
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
JH1
JH2
JH3
JH4
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1
Eμ
DQ52
DSP Utilization
(Number of Cloned Sequences)
20
18
16
14
12
10
8
6
4
2
0
DSP
DPEU
EU+
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5

B

		DSP2.9	Heptamer		Heptamer	JH1
Germline	TACTGTG	TCTATGATGGTTAC	CACAGTG		CACTGTG	CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
		DSP2.9 to JH1		**(N)**		
EU+_DSP2.9	TACTGTG	TCTATGATGTTACTAC		GCCTTCA		CTACTGGTACTTCGATGCTGGGGCGCAGGGACCACGGCACCGTCTCCTCA
DPEU_DSP2.9	TACTGTG	TCTATGATGGTTACTAC		GTGACCA		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.9	TACTGTG	TCTATATGGTTACTAC		GCGAGAG		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA

		DSP2.9	Heptamer		Heptamer	JH2
Germline	TACTGTG	TCTATGATGGTTAC	CACAGTG		CACTGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
		DSP2.9 to JH2		**(N)**		
EU+_DSP2.9	TACTGTG	TCTATGATGGTTACTAC		TCCTACT		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.9	TACTGTG	TCTATGATGGTTACTA		GAGCTAACTCC		TTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA

		DSP2.9	Heptamer		Heptamer	JH3
Germline	TACTGTG	TCTATGATGGTTAC	CACAGTG		CACTGTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
		DSP2.9 to JH3		**(N)**		
EU+_DSP2.9	TACTGTG	TCTATGATGGTTACT		ACGATGAAA		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.9	TACTGTG	TCTATGATGGTTACTAC		-		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA

		DSP2.2	Heptamer		Heptamer	JH1
Germline	TACTGTG	TCTACTATGATTACGAC	CACAGTG		CACTGTG	CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
		DSP2.2 to JH1		**(N)**		
EU+_DSP2.2	TACTGTG	TCTACTATGATTACG		CCTGGTT		GGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		ACCTTTG		ACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATACGAC		-		CTACTGGTACTTCGATGTCGGGGGCAGGGACCACGGTCACGTCTCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		TGGCGGC		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.2	TACTGTG	TCTACTATGATTACGAC		GGAACAGTGAC		CTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA

		DSP2.2	Heptamer		Heptamer	JH2
Germline	TACTGTG	TCTACTATGATTACGAC	CACAGTG		CACTGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
		DSP2.2 to JH2		**(N)**		
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		-		TTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		TGGGGCC		ACTTTGACTACTGGGCCAAGGCACACTCTCACAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACG		GGGC		TACTTTGACTACTGGGGCCAGGCACCACTCTCACAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		TTGGCCCC		TACTTTGACTACTGGGGCCAGGCACCACTCTCACAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		GGATTCTGG		TTGACTACTGGGCCAAGGCACCACTCTCAAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		AAGCTTC		TTGACTACTGGCCAAGCACCACTCTCAAGTCTCCTCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		ACTACTGGG		TACTTTGACTACTGCCAAGGCACCACTCTCACAGTCTCCTCA

		DSP2.2	Heptamer		Heptamer	JH3
Germline	TACTGTG	TCTACTATGATTACGAC	CACAGTG		CACTGTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
		DSP2.2 to JH3		**(N)**		
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		CAGGATGA		GCCTGGTTGCTTACTGGGGCAAGGGACTCTGGTCACTTCTCTGCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGA		TTAGAA		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTC
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		GGACGGTT		CCTGGTTGCTTACTGGGCCAAGGACTCTGGTCACTGTCTCTGCA
EU+_DSP2.2	TACTGTG	TCTACATGATACGAC		GACGCAGA		GCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
EU+_DSP2.2	TACTGTG	TCTACTATGATTACGAC		TCAGGGA		GCTGTTTGCTTATGGGGCAAGGGACTCTGTCACTGTCTCTGCA

D

	DSP2.x	Heptamer		Heptamer	JH2
Germline	TACTGTG	CCTACTATAGTAACTAC	CACAGTG	CACTGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
		DSP2.x to JH2		(N)	
EU+_DSP2.x	TACTGTG	CCTACTATAGTAACTAC		GCTGGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.x	TACTGTG	CCTACTATAGTAACTAC		AGATCTTGCA	TTGACTACTGGGGCCAAGCACCACTCTCACAGTCTCCTCA

	DSP2.x	Heptamer		Heptamer	JH3
Germline	TACTGTG	CCTACTATAGTAACTAC	CACAGTG	CACTGTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
		DSP2.x to JH3		(N)	
EU+_DSP2.x	TACTGTG	CCTATATAGTAACTAC		GATTCTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA

E

		DSP2.3/4	Heptamer		Heptamer	JH1
Germline	TACTGTG	TCTACTATGGTTACGAC	CACAGTG		CACTGTG	CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
		DSP2.3/4 to JH1		**(N)**		
EU+_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		TACCG		CTACTGGTACTCGATGTTGGGGCGCAGGGCCACGTCACCGTCTCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		GAGCTTGG		CTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		CAGGTAAGCTG		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		CGGAAGCAT		CTACTGGTACTTCATGTCTGGGCGCAGGGACCACGGTACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGCAGCTA		CTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		GCTACTGGTACT		TGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		ATAGATTTTACTGC		CTACTGGACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		GCCGGAAGCATAAGT		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGGTCATGTAGAC		GTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTCA		CCGGACAGCA		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGGGGTGGG		TACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCT
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTAC		AAAGAACA		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		CGACGGG		CTACTGGTACTTCGATCTGGGGGCAGGGACCACGGTCACCGTCTCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACTAC		TAAAGCCTG		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACTA		GCCGCTAGG		GTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		TCCAT		TACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTA		-		GGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTAC		CAGAGTAGCATG		GTACTTCGATGTCTGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGA		AAGCCTGG		TGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		CGTGCCAGC		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		CCTGCTA		GTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA

F

		DSP2.3/4	Heptamer		Heptamer	JH2
Germline	TACTGTG	TCTACTATGGTTACGAC	CACAGTG		CACTGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
		DSP2.3/4 to JH2		**(N)**		
EU+_DSP2.3/4	TACTGTG	TCTACTATGTTACGAC		AGGGCTC		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGTAACA		TTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GCGGTCCCG		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		-		ACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGATACGA		CACAGTAGTAGAT		TACTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCT
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		ACCAAGATTCTC		TTGACTACTGGGGCCAAGGCACCATCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGA		GGTGA		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTAC		GGCGTAAG		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGA		TGGCAGCTTG		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTAC		CTTGGTCCT		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTAC		TCCCCTC		TACTTTGACTACTGGGCCAAGGCACCACTCTCACAGTCTCTCA
DPEU_DSP2.3/4	TACTGTG	TCTATGATGGTTACTAC		TCTGAGCT		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACT		GTGAGTCC		ACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		-		TTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GTCCTTG		CTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		GGTAGCAGT		GACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		TTGGG		TACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA

		DSP2.3/4	Heptamer		Heptamer	JH3
Germline	TACTGTG	TCTACTATGGTTACGAC	CACAGTG		CACTGTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
		DSP2.3/4 to JH3		**(N)**		
EU+_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGGGCTA		GCCTGGTTTCTTACTGGGGCCAAGGGACTCTGGTTGTCTCTGCA
EU+_DSP2.3/4	TACTGTG	TCTACTATGTTACGA		GTCCTTC		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GCCGGAAGC		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGA		GTGCCTAGA		TGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GCGGCTT		TGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.3/4	TACTGTG	TCTACTCCGTACTAG		-		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		AGGACTTCTCC		TTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACTAC		GGAAGCATA		GCCTGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTA		GGTAG		GTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		GGGTTGA		GCCTGGTTGCTACTGGGCCAGGGACTCTGGTACTGTCTCTGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GACCTGG		TGGTTTGCTTACTGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACG		AGTGTCAT		TTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DPEU_DSP2.3/4	TACTGTG	TCTACTATGGTTACGAC		GGGGGC		TGGTTTGCTTACTGGGGCCAAGGGACTCTGTCACTGTCTCT

H

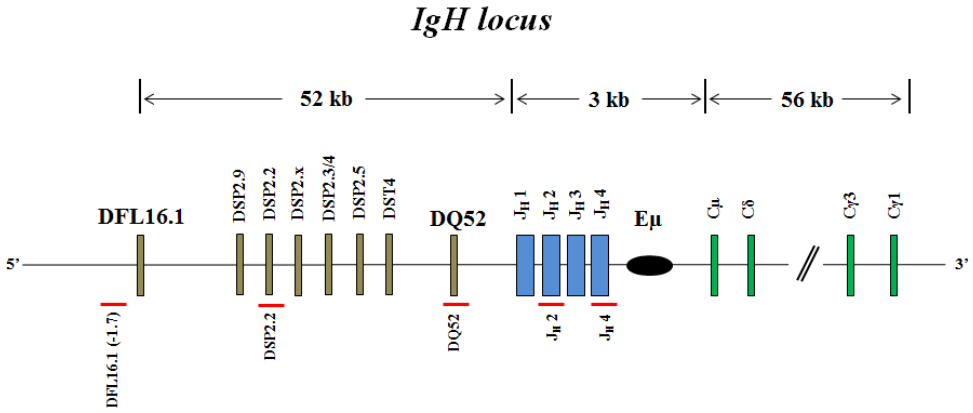
		DSP2.5	Heptamer		Heptamer	JH1
Germline	TACTGTG	TCTACTATGGTAACTAC	CACAGTG		CACTGTG	CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
		DSP2.5 to JH1		**(N)**		
EU+_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		GTGGCCT		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
EU+_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		CCTTTGA		CTACTGGTACTGCGATGTCTGTGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.5	TACTGTG	TCTACTATGGTAACTA		ATTCTGAAC		CTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DSP_DSP2.5	TACTGTG	TCTACTATGGTAACTA		TGAGTCCT		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.5	TACTGTG	TCTACTATGGTAACT		-		CTACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
DPEU_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		TGTCGTGC		GTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA

		DSP2.5	Heptamer		Heptamer	JH2
Germline	TACTGTG	TCTACTATGGTAACTAC	CACAGTG		CACTGTG	TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
		DSP2.5 to JH2		**(N)**		
EU+_DSP2.5	TACTGTG	TCTACTATGGTAACT		CCGTGG		TCTTGACTACTGGGGCAAGGCACCACTCTCACAGTCTCCTC
EU+_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		GTCACTA		TACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCA
DSP_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		GGATTCTGGG		TTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTC

		DSP2.5	Heptamer		Heptamer	JH3
Germline	TACTGTG	TCTACTATGGTAACTAC	CACAGTG		CACTGTG	GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
		DSP2.5 to JH3		**(N)**		
EU+_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		GGATTCTG		GCCTGGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DSP_DSP2.5	TACTGTG	TCTACTATGGTAACTAC		GGCTAC		GGTTTGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA
DPEU_DSP2.5	TACTGTG	TCTACTATGGTAACT		GGGGCTTAC		TGCTTACTGGGGCCAAGGGACTCTGGTCACTGTCTCTGCA

Figure 16. Eμ Induced D-gene Rearrangements in Transgenes. DSP-JH rearrangement PCR products were fractionated through a 1% agarose gel, extracted, and bands for JH1, JH2, and JH3 were cloned into a TA Cloning Vector (Invitrogen, catalog # K2020-20) for EU+, DSP, and DPEU transgenic strains. Clones were then sequenced and the number of DH¬ genes utilized for each transgenic strain is given in (A). A table of the sequences is given in (B - F) with the germline sequence at top and each DH to JH rearrangement underneath. A list for DSP2.9 is in (B), DSP2.2 in (C), DSP2.x (D), DSP2.3/4 in (E - G), and DSP2.5 in (H).

Figure 17


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
Jμ1
Jμ2
Jμ3
Jμ4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-1.7)
DSP2.2
DQ52
Jμ2
Jμ4

A

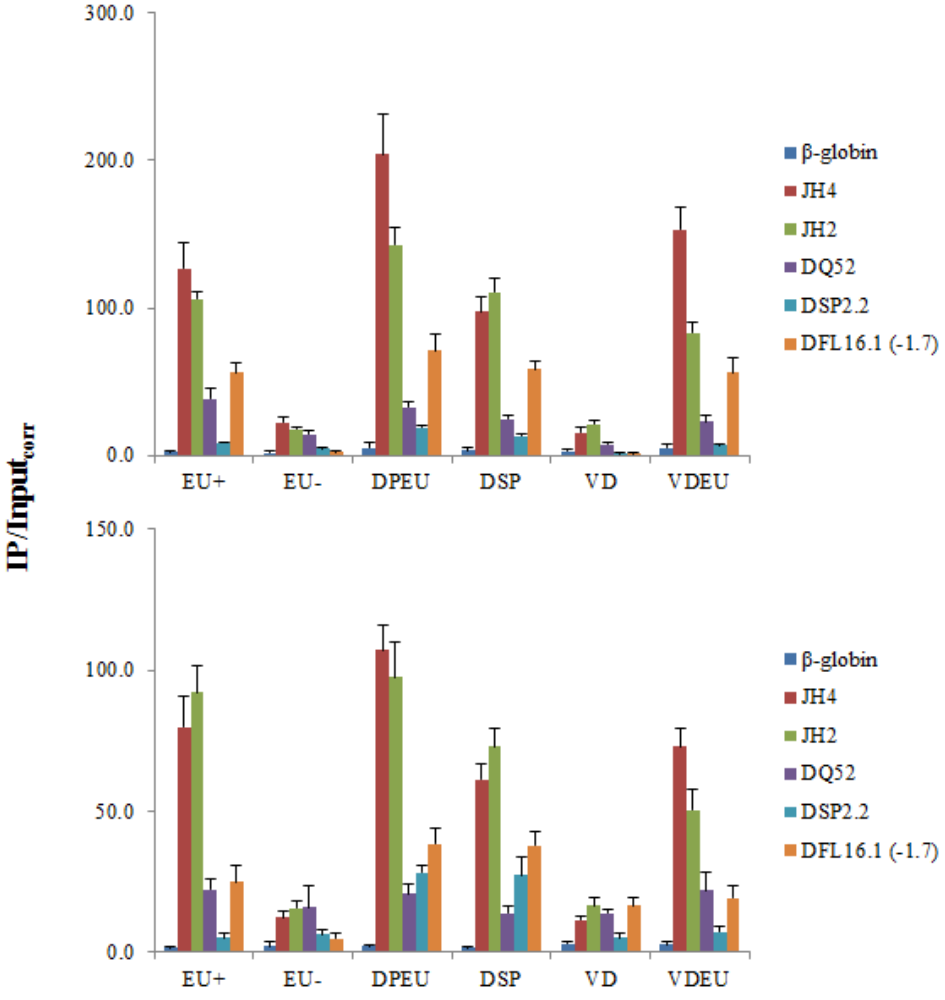
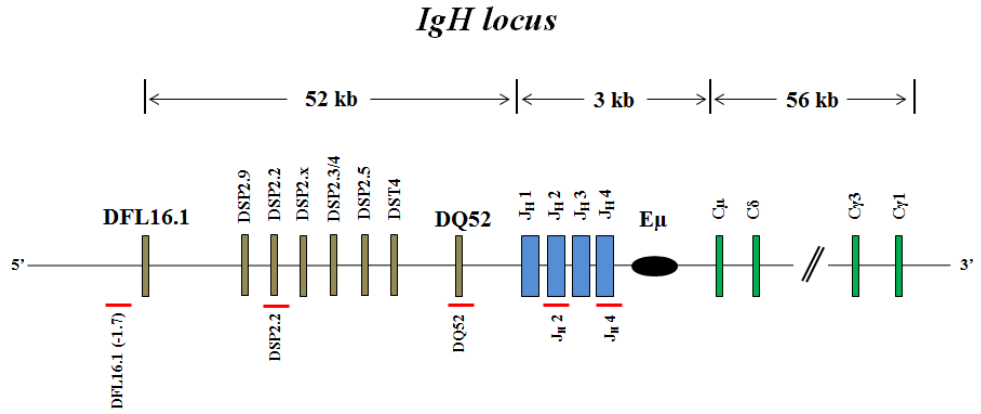

IP/Input corr
300.0
200.0
100.0
0.0
EU+
EU-
DPEU
DSP
VD
VDEU
β-globin
JH4
JH2
DQ52
DSP2.2
DFL16.1 (-1.7)
150.0
100.0
50.0
0.0
EU+
EU-
DPEU
DSP
VD
VDEU
β-globin
JH4
JH2
DQ52
DSP2.2
DFL16.1 (-1.7)

Figure 17


IgH locus
52 kb
3 kb
56 kb
DFL16.1
DSP2.9
DSP2.2
DSP2.x
DSP2.3/4
DSP2.5
DST4
DQ52
JH1
JH2
JH3
JH4
Eμ
Cμ
Cδ
Cγ3
Cγ1
5'
3'
DFL16.1 (-1.7)
DSP2.2
DQ52
JH 2
JH 4

B

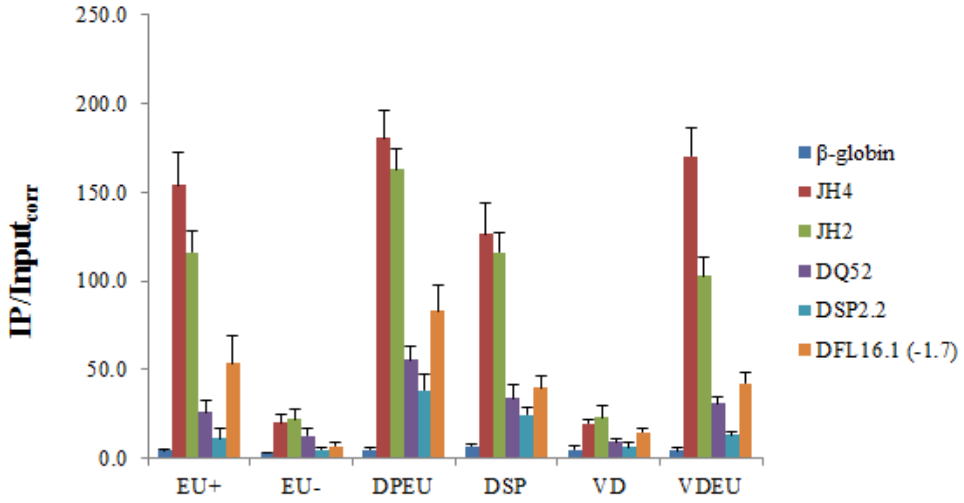

IP/Input corr
250.0
200.0
150.0
100.0
50.0
0.0
EU+
EU-
DPEU
DSP
VD
VDEU
β-globin
JH4
JH2
DQ52
DSP2.2
DFL16.1 (-1.7)

Figure 17. Binding of RAG-1 and RAG-2 in Transgenic Mice. Binding of RAG-1 (A) or RAG-2 (B) was assessed by ChIP using primary CD19+ pro-B cells from the bone marrow of representative transgenic mice. RAG-1 or RAG-2 specific antibody was used, followed by qPCR with 250 pg DNA in duplicates. Amount of RAG-1 or RAG-2 protein at different gene locations ((IP/Inputcorr) was calculated as described (Ji et al., 2010). Primers for amplification are indicated by the red bars in the schematic above each figure. Data is from two independent experiments for each founder. Founder set 1 is on top and founder set 2 is on bottom for (A). Only data for founder set 1 was generated for the RAG-2 ChIP in (B). Immunoprecipitates were analyzed in duplicates. Error bars represent standard deviation between experiments.

Chapter 3

Genome-Wide DNase I Hypersensitivity Study Of The Eμ Enhancer On Mouse Chromosome 12

Introduction

Gene regulatory sequences that exert developmental and tissue-specific control often undergo changes in DNase I-sensitivity that parallel transcription factor occupancy (Lazarovici et al., 2013). Such regulatory genes include the mouse proto-oncogenes, *c-fos* and *c-myc*, whose DNase I-sensitivity changes as a result of transcription induction (Chen et al., 1987; Chen et al., 1990). These active regulatory sequences are associated with the presence of DNase I hypersensitive sites in the genome since DNA binding at these sites is more accessible to transcription factors and accessory proteins due to their nucleosome depletion (Gross and Garrard, 1988). The location of DNase I hypersensitive sites (DHSs) in a genome is then critical to understanding where regions of open chromatin exist that regulate gene expression in a developmental and tissue-specific manner (Sheffield et al., 2013).

Several DHSs have been previously identified within the immunoglobulin (*IgH*) locus of mouse chromosome 12. At the 5' end of the locus, a cluster of four DHS peaks near Zfp386, a Kruppel-like zinc finger protein, were located 30 kb upstream of the last 5' V_H gene (Pawlitzky et al., 2006). Further analysis on these sites revealed no relationship to allelic exclusion or gene recombination (Matheson and Corcoran, 2012). In the 100 kb V_H to D_H intergenic region, which is between the 3' most V_H gene (V7183.1.1pg) and DFL16.1, six hypersensitive peaks were identified where three of the sites were shown to have an insulator sequence that contains CTCF-binding elements (Featherstone et al., 2010). It was proposed that the function of these CTCF sites serve as a barrier element to recombination and prevent spurious germline rearrangements between V_H, D_H, and J_H gene segments. Within the D_H gene segment, the intervening D_H

genes are insensitive, while the outer 5' DFL16.1 and 3' DQ52 have been reported to be mildly sensitive and hypersensitive to DNase I, respectively (Chakraborty et al., 2009). Three additional DHS clusters have been identified that are important in the *IgH* locus. This includes the four sensitive J_H genes that are ~1 kb 3' of DQ52 (Maës et al., 2006), as well as a cluster of hypersensitive peaks in the 3' region of the *IgH* locus that led to the identification of the 3' regulatory region (3'RR) (Giannini et al., 1993; Michaelson et al., 1995). This 3'RR is critical for somatic hypermutation and class switch recombination, but not for V(D)J joining (Rouaud et al., 2012).

The only DNase I hypersensitive site on the *IgH* locus that has been demonstrated to have a significant role in both epigenetic modifications and V(D)J recombination is the 700 bp Eμ enhancer located in the intronic region between J_H4 and Cμ (Chakraborty et al., 2009; Mills et al., 1983). Whereas the Eμ enhancer is critical for activating *IgH* transcription and recombination, it remains unknown whether it is also involved in regulating genes distant from the *IgH* locus, as well as those located on other chromosomes. Earlier studies demonstrated that DHS associated with the DQ52 promoter was lost upon deletion of the Eμ enhancer, which has been attributed to the promoter becoming inactive by loss of the Eμ enhancer (Chakraborty et al., 2009). Based on this observation, we reasoned that one way to search for Eμ-mediated interactions elsewhere in the *IgH* locus and on other chromosomes was to screen genome-wide for DHS sites that are lost upon deletion of the Eμ enhancer. We used genome-wide DHS sequencing to probe the open chromatin structure at the single nucleotide level to determine whether the Eμ enhancer in mouse pro-B cells affects other genes on chromosome 12 or elsewhere in the murine genome. Our results show unique DHS

peaks in some areas of the *IgH* locus in both the presence and absence of the Eμ enhancer. In addition, our preliminary analysis did not detect the Eμ enhancer having an effect on other genes in other chromosomes, although this remains to be completed using additional computational resources for processing the DNase I-seq data.

Experimental Flow

DNase I-seq is a global and high-resolution method that uses the nonspecific endonuclease DNase I to map chromatin accessibility. These accessible regions are designated as DNase I hypersensitive sites (DHSs) and define sequences of regulatory potential (e.g., promoters, enhancers, insulators, and locus control regions) in complex genomes. In this study, the B6D345, Rag8-/- and Rag2-/- Eμ-/- cell lines were used to screen any changes in DHS peaks genome wide as a result of Eμ enhancer deletion. Both Rag8-/- and Rag2-/- Eμ-/- cell lines are from the 129 mouse strain, while the B6D345 cell line is from the C57BL/6J strain. The B6D345 cell line was included as a wild-type control since it maps directly to the C57BL/6J reference strain on the UCSC mm9 genome browser. A complete description of these cell lines and the DHS experiment is given in the Materials and Methods section.

Briefly, 1 x 10^7 nuclei from cell lines carrying wild-type or Eμ-deleted IgH alleles were digested with a range of concentrations of DNase I (Worthington-Biochem, catalog# DPFF), in order to separate the most accessible chromatin into short fragments. Typically, a range of 0 to 100 units/ml of DNase I concentrations were used per 1 x 10^7 nuclei. One critical aspect that was primary to the success of this experiment was the replacement of Mg+2 with Ca+2 in the digestion buffer to slow the enzymatic activity of DNase I prior to ultracentrifugation (Figure 18). Real-time PCR was also used to determine the optimal level of DNase I enzyme to prevent over-digestion (Figure 19). This was reflected in several ways. First, the real-time PCR results demonstrated little digestion for insensitive genes at all concentrations (NFM, Cy3, β-globin). But, for hypersensitive gene regions (Eμ, β2M, DQ52), there was increasing loss of DNA with

increasing units of DNase I (Figure 19A,B). When Eμ is deleted in the Rag2-/- Eμ-/- cell line, specific *IgH* hypersensitive loci are insensitive, including DQ52 and Eμ (Figure 19C). Second, the parent band in the gel was intact at all concentrations, with minimal smearing underneath at the 20 unit concentration of DNase I (Figure 18b). This is important, as extensive smearing below the parental DNA band is related to overgrown cells, endogenous nuclease activity, or over-digested chromatin. Overall, we found that the optimum amount of DNase I enzyme is at 20 units/ml for 10 million cells, as it produced clear and consistent sequencing tracks on the UCSC genome browser between samples (Figure 21).

The short DNA fragments from the DNase I digestion were then size selected by ultracentrifugation over a sucrose cushion in order to enrich for 200 – 400 bp size fragments (Figure 20). Fractions were then isolated in 500ul volumes and purified. The size-selected DNase I hypersensitive DNA was then prepared for high-throughput sequencing. DNA was ligated with adaptors and amplified to generate a DHS library for single-end sequencing on the Illumina Genome Analyzer (Illumina, catalog # PE-102-1004). We obtained 13.6 to 21.6 x 10^6 non-redundant sequence tags for three samples, repeated two times at DNase I concentrations of 10, 20, and 40 units/ml. Data from the Illumina Genome Analyzer was then analyzed using manufacturer-provided software, uploaded to the UCSC Genome browser, and aligned to the mm9 database (NCBI Build 37) that contains only the C57BL/6J reference strain.

Results and Discussion

First, we found that our genome-wide DHS results were consistent between samples in housekeeping genes that are known to be hypersensitive outside of the *IgH* locus on mouse chromosome 12. In figure 21, β_2-microglobulin, Mb-1, and E2A are all used as examples to demonstrate the consistent peak and patterns that are observed upstream of the respective gene promoters between samples. Also shown in the top track of figure 21 is an outside reference from a DHS experiment using mouse spleen B Cells ($CD43^-$, $CD11b^-$) from the Stamatoyannopoulous Laboratory at the University of Washington.

Second, there was consistency in peaks between different samples across the entire *IgH* locus (chr12: 114,500,000 – 117,500,000) (Figure 22A). There were no significant gaps in the sequence between samples and there was remarkable similarity between samples at both the 5' and 3' ends (Figure 22A). Along with this, there were no discrepancies in mapping the 129 mouse cell line strains ($Rag8^{-/-}$ and $Rag2^{-/-}E\mu^{-/-}$) to the mm9 database that uses the C57BL/6 sequence, as there was exact alignment with the B6D345 cell line that is from the C57BL/6 mouse strain. Moreover, the sequencing readout also clearly showed a peak corresponding to the Eμ enhancer (chr12:114,665,501) in both sets of cell lines that contain wild-type *IgH* alleles, while no peak was present in the same position for the cell line with Eμ - deficient alleles (Figure 22B). These results support the real-time PCR data (Figure 19) and previous data that demonstrated Eμ dependent affects with DQ52 and DFL16.1 (Chakraborty et al., 2009). Specifically, when the Eμ enhancer was present, both Eμ and DQ52 demonstrate strong hypersensitivity in the qPCR (Figure 19A, 19B) and the genome-wide DHS sequencing

data (Figure 22B). When the Eμ enhancer was absent in the Rag2$^{-/-}$ Eμ$^{-/-}$ cell line, there was significant loss of the DQ52 (chr12: 114,668,794) associated DHS signal (Figure 19C) and peak (Figure 22). Likewise, DFL16.1 (chr12: 114,720,334) was moderately sensitive when the Eμ enhancer was present and this peak was reduced when the enhancer is absent (Figures 19C, 22B).

Such consistency in the trends from the genome-wide DHS tracks for both within and outside the *IgH* locus allowed deeper exploration of the data to explore the effects of the Eμ enhancer both within the *IgH* locus and outside of it. Two additional tracks are present in Figure 22 that allows comparison to mouse pro-B cell and spleen B cell DHS peaks. The mouse spleen B cell (CD43^{-}, CD11b^{-}) DHS is from the Stamatoyannopoulous Laboratory at the University of Washington (UW) and the mouse pro-B cell DHS is from the Busslinger laboratory at Vienna Biocenter (VB). In particular is the identification of a unique DHS peak that is approximately 20 kb upstream of the Eμ enhancer at position 114,684,105, between DQ52 (chr12: 114,668,794) and DST4 (chr12:114,686,510). This peak is present in all tracks where the Eμ enhancer is present except for the spleen B cell DHS. When the Eμ enhancer is absent in the Rag2$^{-/-}$ Eμ$^{-/-}$ track (Figure 22B, bottom), the peak at 114,684,105 is reduced to the same level as DQ52. Therefore, this site represents a new area of regulation by the Eμ enhancer and may have a role in D_H to J_H recombination since it is adjacent to the outer DSP gene segment, DST4.

In addition, we identified two other significant DHS-peaks within the *IgH* locus not dependent on the Eμ enhancer. The first peak was between Cγ1 and Cγ2b located at position 114,558,806 - 114,560,261 on mouse chromosome 12 (Figure 23). The peak was significantly stronger than other peaks within the 400 kb window in Figure 23, and

therefore represents a potentially significant cis-acting element. It was also present in the VB data track and was absent in the UW data track. It is of particular importance to note the consistency of the presence of the peak in all three samples (Rag8-/-, B6D345, and Rag2-/- Eμ-/-) (Figure 23). One way to compare the relative strength of DHS peaks in terms of sensitivity is to compare them to known hypersensitive sites. The *IgH* 3'RR has four enhancers (hs3a, hs1.2, hsb, and hs4) and four DNase I hypersensitive sites were identified, known as hs5-8 (Garrett et al., 2005). These sites are located ~25 kb downstream of the most 3' C_H genes, Cα (Pettersson et al., 1990). Since it is known to have strongly hypersensitive sites, comparison of peaks to the 3'RR is one way to assess the relative strength of DHS peaks. This peak between Cγ1 and Cγ2b was present in the Rag8-/-, B6D345, and Rag2-/- Eμ-/- tracks and was similar to the sensitivity in the 3'RR. Yet, in the VB data, the 3'RR DHS peaks were significantly reduced and therefore it is difficult to assess the relative strength of this peak between Cγ1 and Cγ2b in this track. One reason for this difference between data sets is due to the methods used to prepare the DNase I digested DNA for sequencing. As noted in the Materials and Methods section, several steps were implemented into the protocol to prevent over and under-digestion of samples in order to execute the purpose of identifying other potential cis-regulatory elements. Our interpretation is that the VB DHS track was over-digested with DNase I enzyme, as noted for the 3'RR and also for DFL16.1. When genome-wide sequencing DHS is performed on higher units of DNase I enzyme that reflect an overly digested state, which is based on additional gel smearing in the gel bands prior to ultracentrifugation and less than 80% of DNA remaining in the pre-sucrose sample measured by qPCR, the DHS peaks are reduced (data not shown).

The second unique peak identified was located at position 115,850,147 - 115,850,965 on mouse chromosome 12 (Figure 24). This is the same location of J558.8.98, which is a V_H gene segment that is upstream of J558.6.96 (chr12:115,821,780-115,822,073) and downstream of J558.9.99 (chr12:115,854,061-115,854,354). This peak has several features that lend it to appear as a significant cis-regulatory element. First, within the 400 kb window shown from the UCSC genome browser, it is the tallest peak in the region for the Rag8-/-, B6D345, and Rag2-/- Eμ-/- tracks. Second, its average peak height between samples is 6.4, which is comparable to other enhancer regions on the *IgH* locus, including the 3'RR (7.5) and the Eμ enhancer (8.4). While this peak was present in both the UW and VB tracks, it is not the strongest DHS in the 400 kb window for these tracks (Figure 24). This peak was not present in either a fibroblast or embryonic stem cell DHS from ENCODE (data not shown). Therefore, this unique peak in the Rag8-/-, B6D345, and Rag2-/- Eμ-/- tracks is specific for pro-B cells and may have be involved in positioning the *IgH* locus for V(D)J rearrangement.

Lastly, there have been previous reports of other DHS peaks on the *IgH* locus (Figure 25). This genome-wide DHS study allowed comparison between DHS sensitivity peaks across the *IgH* locus. A specific window of the 3'RR is given in figure 25A to assess the other areas of the *IgH* locus. There have been six DHSs identified within the 100 kb intergenic $V_H - D_H$ region (Featherstone et al., 2010), and four DHS peaks 30 kb upstream of the last V_H gene, J558.89pg.195 (chr12: 117,263,407) (Pawlitzky et al., 2006) (Figure 25). In the present genome-wide DHS-seq analysis, only HS4 was identified within a 10 kb region upstream of DFL16.1 that is within the 100 kb intergenic $V_H - D_H$ region (Figure 25). Every other peak within the region was of only weak to

moderate sensitivity. One significant peak was detected 30 kb upstream of the last V_H gene near Zfp286, which is referred to as Hs3b in the Pawlitzky study, but all other peaks in this region were also only moderate to weak in DNase I sensitivity (Figure 25). Therefore, one distinct advantage of this genome-wide DHS study in pro-B cells is the global view afforded to the *IgH* locus to interrogate all significant DHS peaks. Not all reported peaks are strong hypersensitive sites, and only a limited number of peaks appear significant on the *IgH* locus. The exact significance of weak and moderate hypersensitive sites needs further exploration.

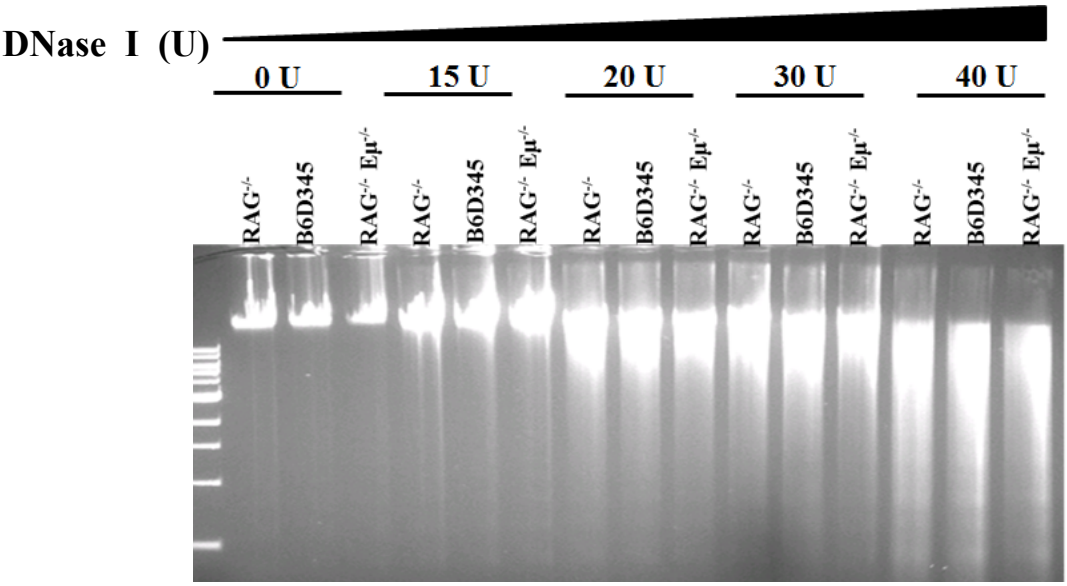
Furthermore, more in-depth bioinformatic analysis on these sequencing results will take place to identify other potential regulatory elements. Using excel-based algorithms, there were no detectable regions in the genome that appeared to be affected by the loss of the Eμ enhancer in the Rag2-/- Eμ-/-. This does not exclude the possibility of a *trans* regulatory activity for the Eμ enhancer, but finer techniques to identify DHS hotspots as a result of Eμ loss will need to be applied to determine the extent of the Eμ to other regions of the genome.

We reached several significant conclusions based on this this finely tuned DHS experiment and results. First, the genome-wide mapping of DNase I hypersensitive (DHS) sites in mouse pro-B cells has revealed novel hypersensitive sites within the *IgH* locus. One significant peak was identified in this study that appears to be regulated by the Eμ enhancer. Due to the peak's position near DST4, this intergenic region likely has a role in D_H to J_H gene rearrangement. Two other significant peaks were identified in this study, but they were present even in the absence of the Eμ enhancer. Therefore, while the Eμ enhancer directly affects some regions of the *IgH* locus, such as DQ52, it does not

affect other regions that appear to be significant. The function of these novel peaks needs to be further explored, as they may have a role in pro-B cell development. This includes positioning the locus for V(D)J rearrangement and even other mechanisms, such as allelic exclusion and class-switch recombination. Second, there are different levels of DHS sites, from strongly hypersensitive to weakly hypersensitive that can only be viewed across the genome in a DHS-seq analysis. While several labs have reported DHS sites in general terms for the *IgH* locus, the exact strength of each DHS site can only be viewed when the whole locus is sequenced. Different laboratories have reported many DHS sites along the *IgH* locus (Figure 5), but only when these sites are viewed as a whole can the strength of each individual peak be ascertained. Many of the DHS sites reported on the *IgH* locus, especially in the intergenic V_H - D_H region, are weak sites that likely serve no regulatory function. On the other hand, several sites, including those identified in this study, have a strong DHS signal, and likely serve a gene regulatory purpose for the *IgH* locus that needs further exploration.

Figure 18

A



B

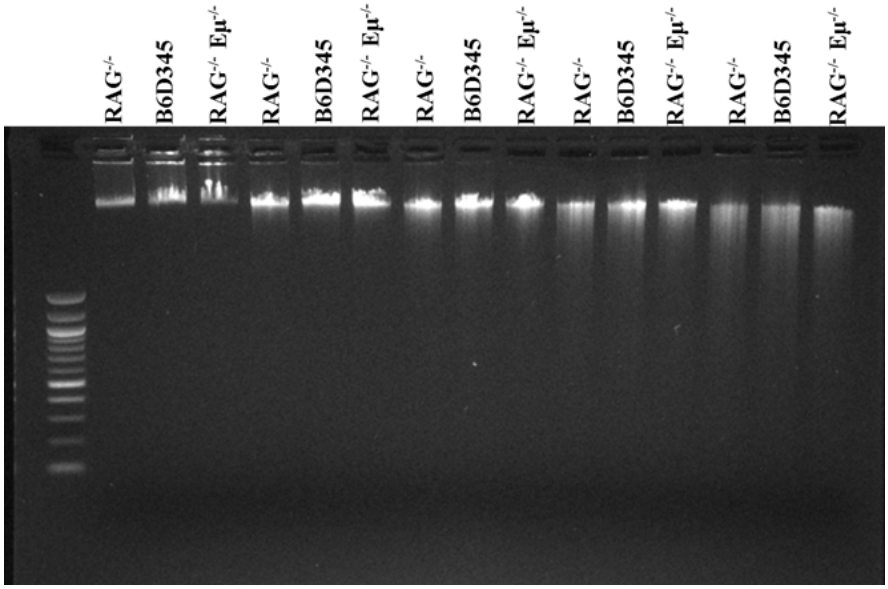


Figure 18. Pre-sucrose DHS gel to Validate DNase I Digestion. Nuclei from DNase I-digested samples were warmed to 55^0C and 3 ul from each of the digested samples were loaded onto a 1% agarose gel and set to run for 45 minutes at 120V. Comparison between over-digested DNA (A) and properly digested DNA (B). The difference between the experiments is the replacement of Mg^{+2} with Ca^{+2} in the digestion buffer to slow down the activity of the DNase I enzyme. When Ca^{+2} is used during the digestion step, a strong and distinct parent band is observed across each digested sample, whereas large smearing and loss of the parent band indicates a poor sample (A). Smearing may also be attributed to overgrown cells, poor nuclei isolation, or incorrect buffer.

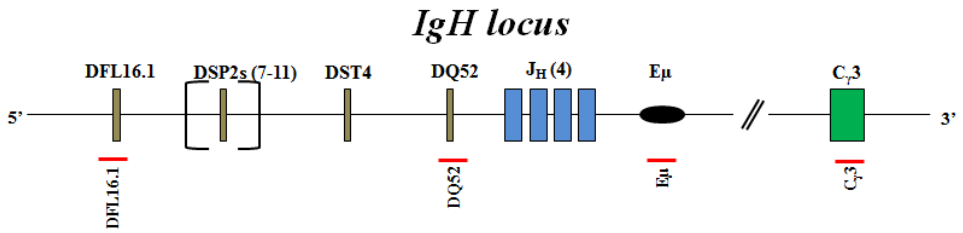
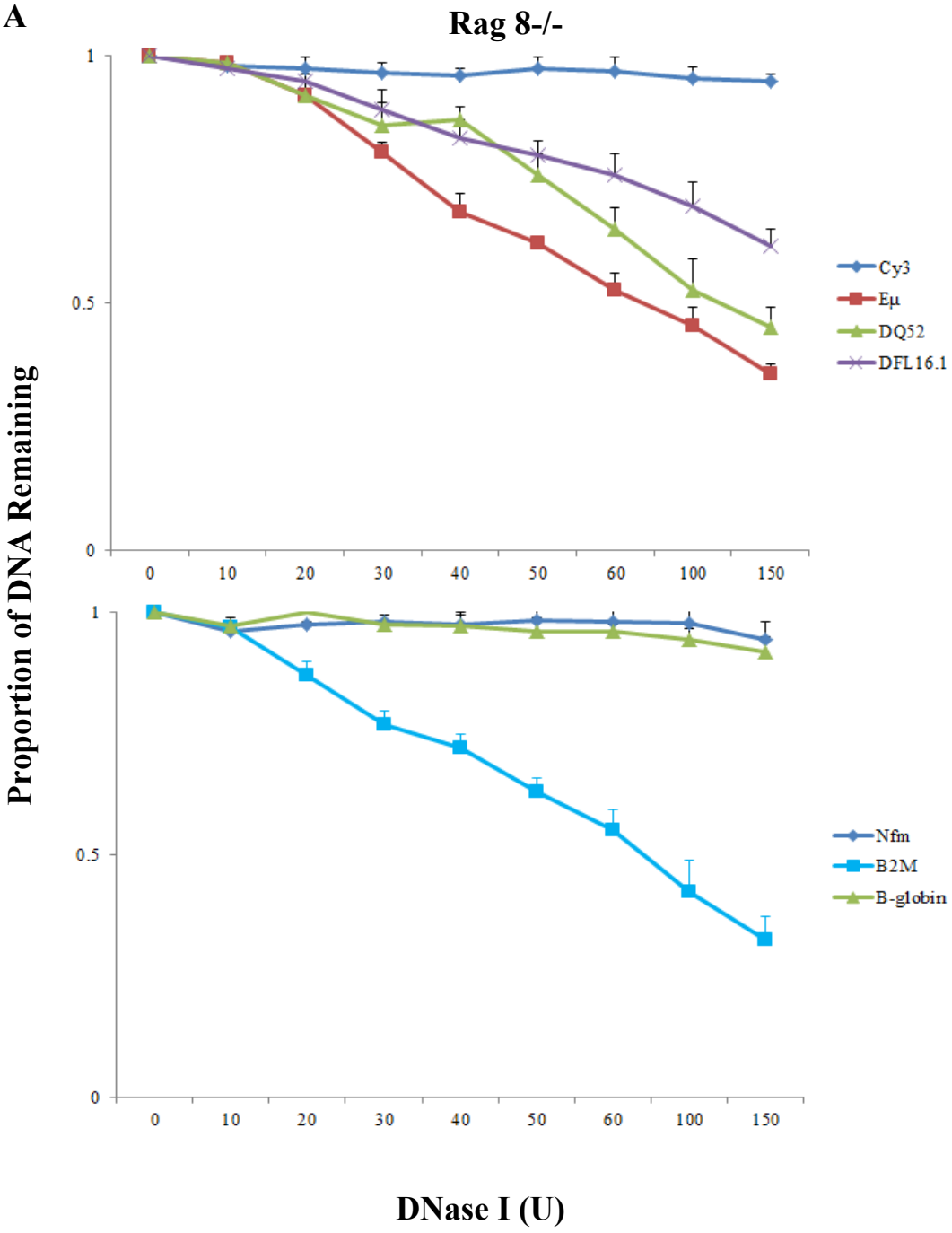

IgH locus
DFL16.1
DSP2S (7-11)
DST4
DQ52
JH (4)
Eμ
Cγ3
5'
3'
DFL16.1
DQ52
Eμ
Cγ3

Figure 19

A


Rag 8-/-
Proportion of DNA Remaining
1
0.5
0
0
10
20
30
40
50
60
100
150
Cy3
Eμ
DQ52
DFL16.1
Nfm
B2M
B-globin
DNase I (U)

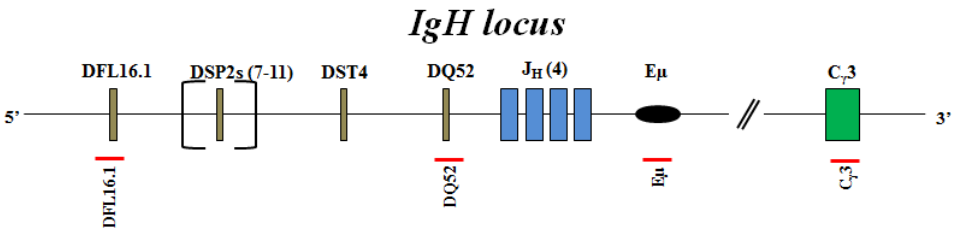
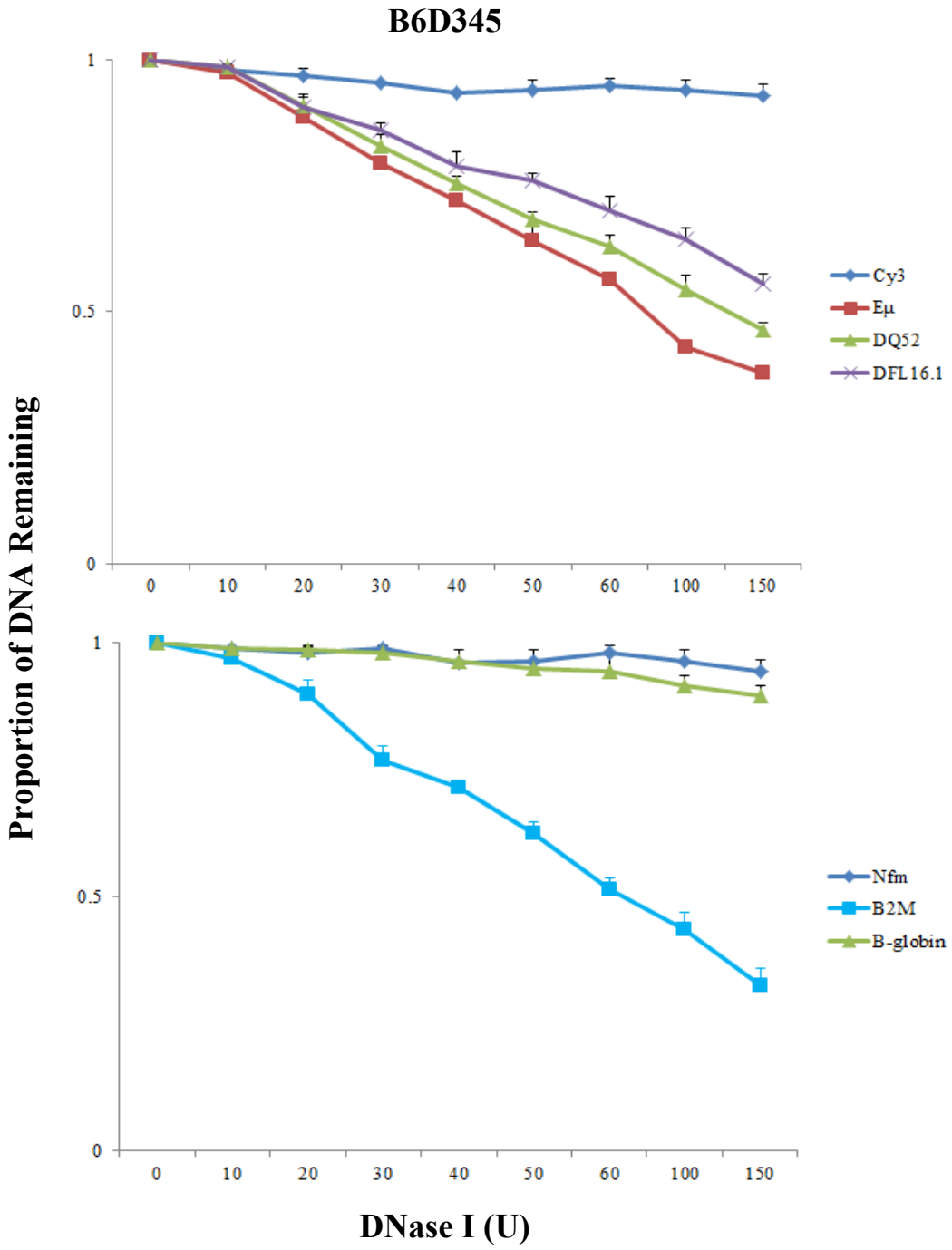

IgH locus
DFL16.1
DSP2s (7-11)
DST4
DQ52
JH (4)
Eμ
Cγ3
5'
3'
DFL16.1
DQ52
Eμ
Cγ3

Figure 19

B


B6D345
Proportion of DNA Remaining
DNase I (U)
Cy3
Eμ
DQ52
DFL16.1
Nfm
B2M
B-globin
0
0.5
1
0
10
20
30
40
50
60
100
150

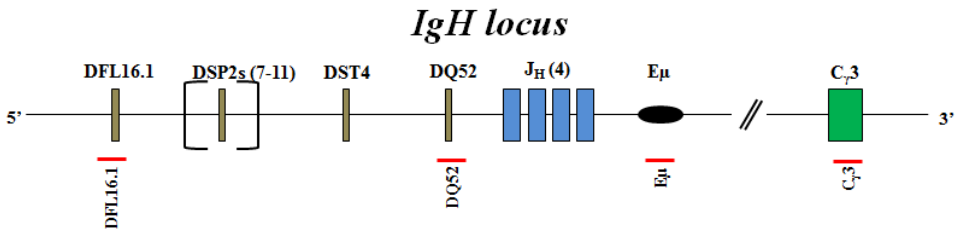

IgH locus
DFL16.1
DSP2s (7-11)
DST4
DQ52
JH (4)
Eμ
Cγ3
5'
3'
DFL16.1
DQ52
Eμ
Cγ3

Figure 19

C

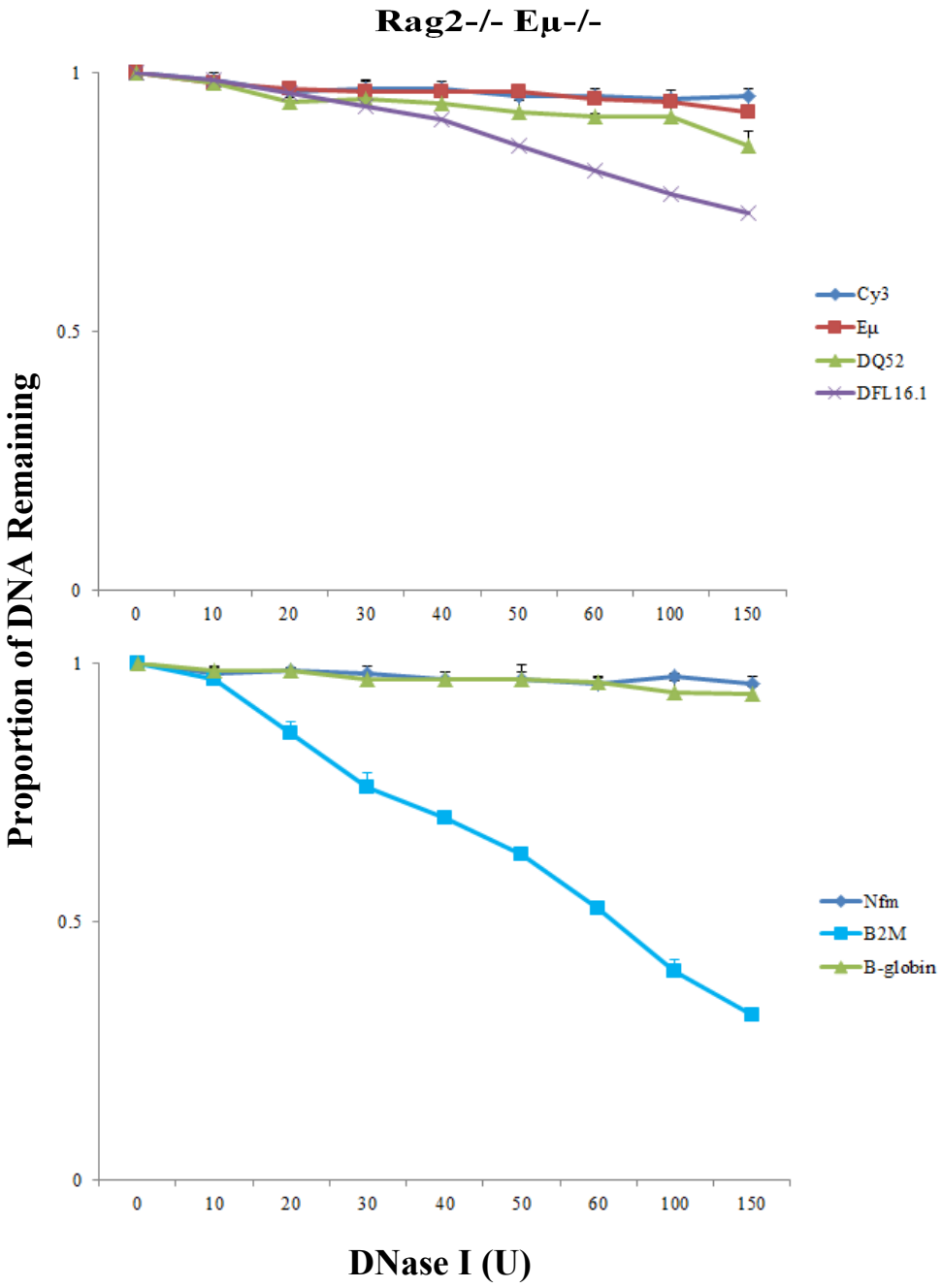

Rag2-/- Eμ-/-
Proportion of DNA Remaining
1
0.5
0
0
10
20
30
40
50
60
100
150
Cy3
Eμ
DQ52
DFL16.1
Nfm
B2M
B-globin
DNase I (U)

Figure 19. qPCR Validation of DNase I sensitivity. Nuclei from each cell line (A) $Rag8^{-/-}$, (B) B6D345, and (C) $Rag2^{-/-}E\mu^{-/-}$ was digested with different concentrations of DNase I (X-axis), and the DNA was then purified from 5 x 10^5 nuclei and a total of 2 ng used for qPCR. Primers were used for both inside the *IgH* locus (top graph) and outside (bottom graph). Primer locations for inside the *IgH* locus are indicated by the black bars underneath each schematic. Primers for outside the *IgH* locus include the following: NFM (chr14:68,741,992), β-globin (chr7:111,010,093) and β_2M (ch2:121,973,178). The proportion of DNA remaining (Y-axis) was normalized to NFM (top graph) or Cγ3 (bottom graph), as described in the Materials and Methods section. Increased DNase I sensitivity is demonstrated with increasing loss of signal as the concentration of DNase I increases. Results are shown for two independent DNase I digestion experiments and error bars represent the standard deviation between experiments.

Figure 20

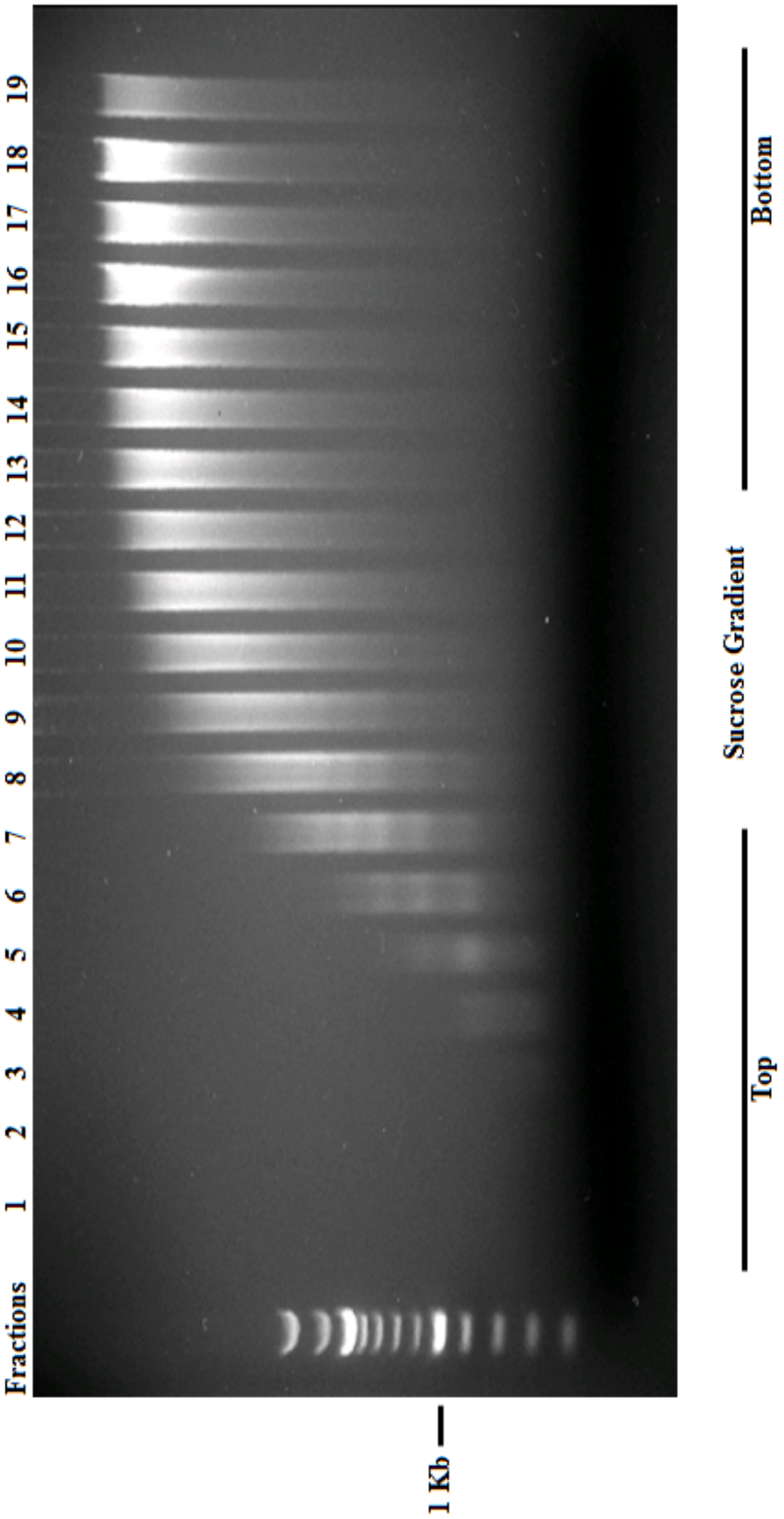
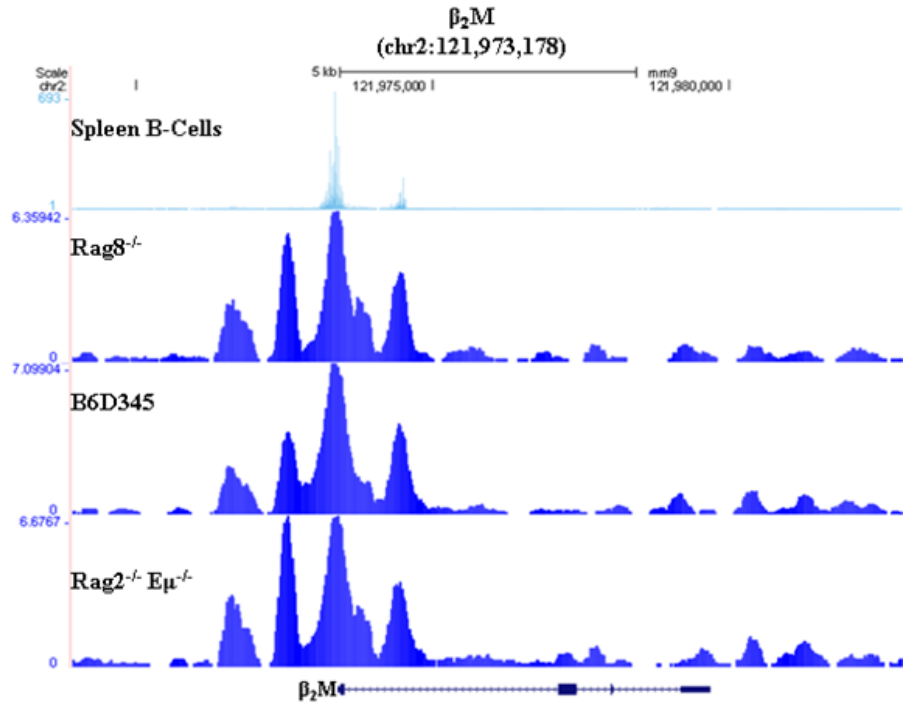

Fractions
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Top
Sucrose Gradient
Bottom
1 Kb

Figure 20. Size Fractionation of DNase I Released Fragments. Nuclei from each cell line was isolated and subsequently subjected to limited DNase I digestion. Small DNA fragments containing DNase Hypersensitive Sites were purified by size selection over a sucrose gradient and 500 ul fractions were collected. The top of the gradient contains the small fragments (<600 bp) (lower left side of gel), while the bottom of the gradient contains the high molecular weight fractions (upper right side of gel). Only those fractions less than 600 bp were used for downstream sequencing. Representative gel is from $Rag8^{-/-}$.

Figure 21

A


β2M
(chr2:121,973,178)
Scale
chr2:
693
5 kb
121,975,000
mm9
121,980,000
Spleen B-Cells
1
6.35942
Rag8-/-
0
7.09904
B6D345
0
6.6767
Rag2-/- Eμ-/-
0
β2M

B

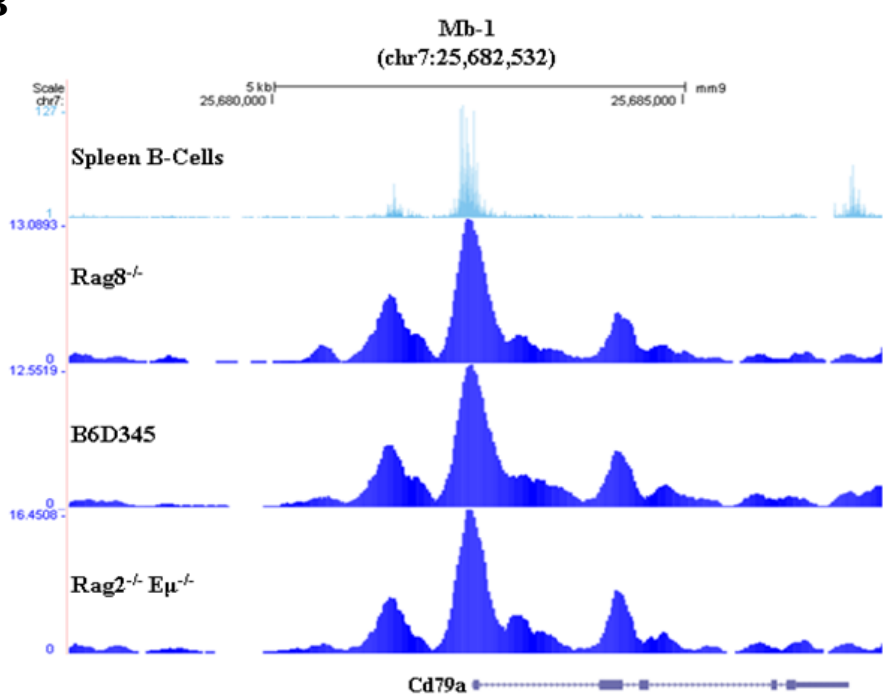

Mb-1
(chr7:25,682,532)
Scale
chr7:
127
5 kb
25,680,000
mm9
25,685,000
Spleen B-Cells
1
13.0893
Rag8-/-
0
12.5519
B6D345
0
16.4508
Rag2-/- Eμ-/-
0
Cd79a

Figure 21

C

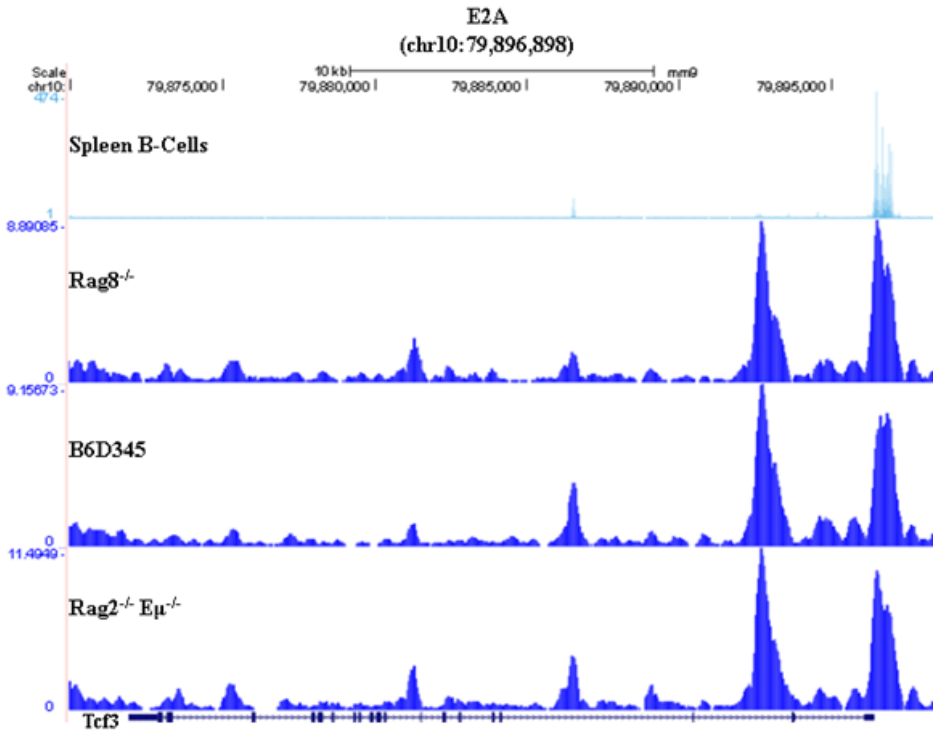

E2A
(chr10:79,896,898)
Scale
chr10:
10 kb
mm9
79,875,000
79,880,000
79,885,000
79,890,000
79,895,000
474
Spleen B-Cells
1
8.89085
Rag8-/-
0
9.15673
B6D345
0
11.4949
Rag2-/- Eμ-/-
0
Tcf3

Figure 21. Consistent DHS Peaks In The Genome Outside of the *IgH* locus. High-throughput sequencing was carried out to obtain 40-nucleotide sequence tags, which were mapped back to the genome and uploaded to the UCSC genome browser to visualize clusters of sequence tags that indicate a DHS site. Four tracks are shown in the UCSC Genome Browser in each figure above for DHS sites that correspond to (A) β_2M (ch2:121,973,178), (B) Mb-1 (chr7:25,682,532), and (C) E2A (chr10:79,896,898). In each figure, the top track is Mouse Spleen B Cells ($CD43^-$, $CD11b^-$) from the Stamatoyannopoulous Laboratory at the University of Washington (GEO dataset: GSM1003813); the next track is $Rag8^{-/-}$, followed by the B6D345 control in the subsequent track, and $Rag2^{-/-}E\mu^{-/-}$ in the bottom track. Results are representative from two independent DNase I digestion experiments for 20 units of DNase I digestion. Enrichment on the Y-axis is measured by the number of sequence tags per million, except for the outside reference track, Spleen B Cells.

Figure 22

A

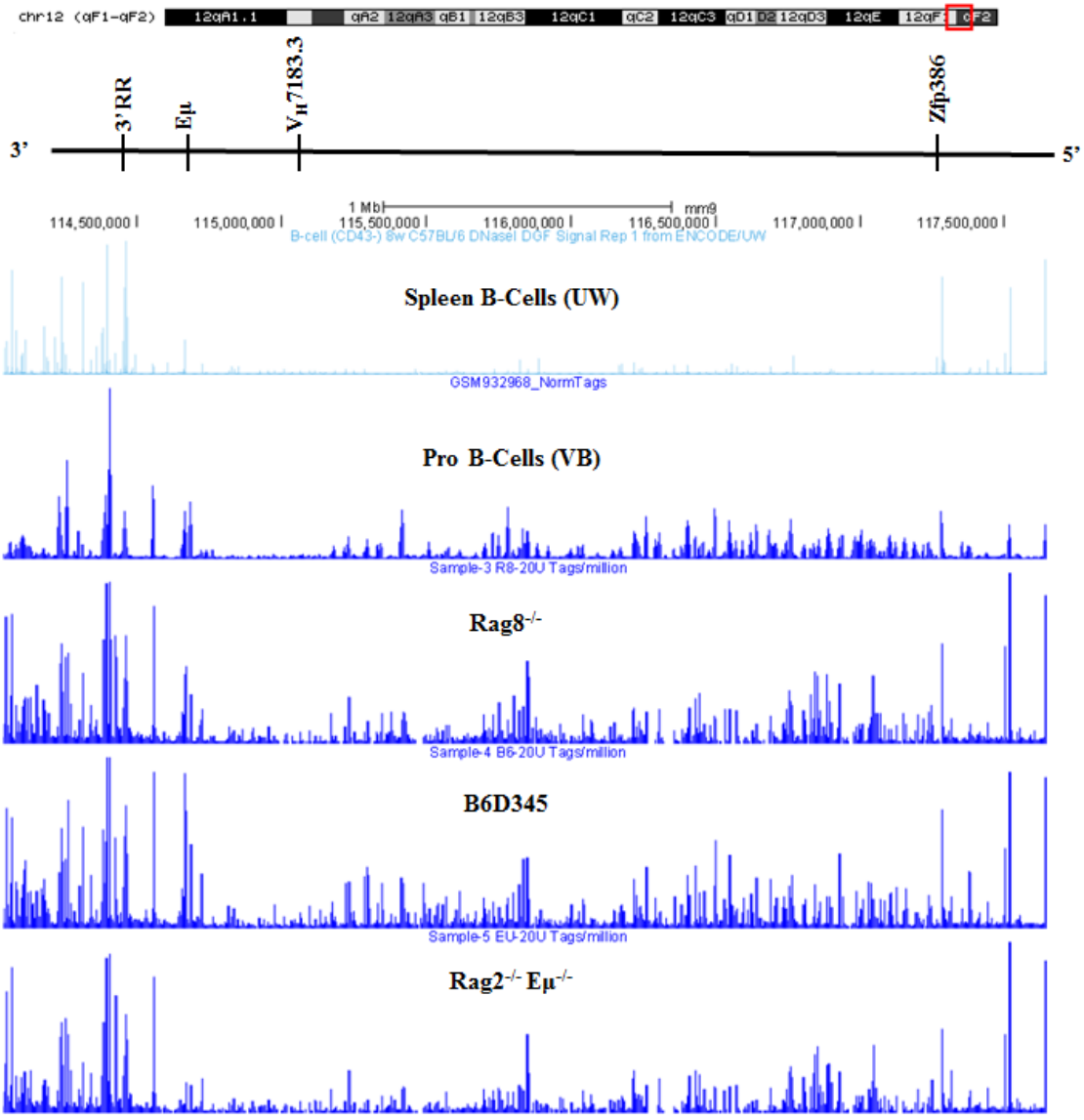

chr12 (qF1-qF2)
3'RR
Eμ
VH7183.3
Zfp386
3'
5'
1 Mb
mm9
114,500,000
115,000,000
115,500,000
116,000,000
116,500,000
117,000,000
117,500,000
B-cell (CD43-) 8w C57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-

Figure 22

B

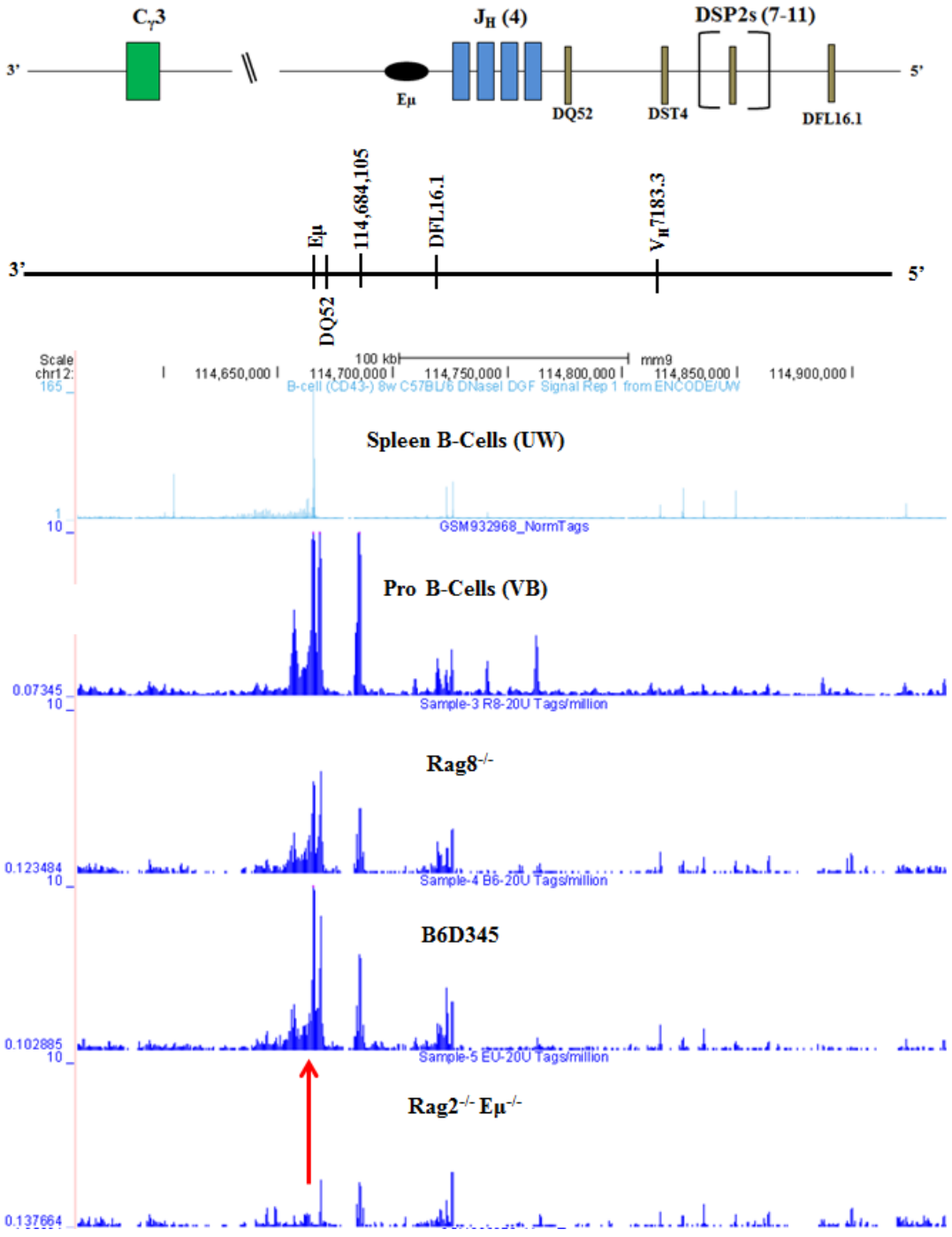

Cγ3
JH (4)
DSP2s (7-11)
3'
5'
Eμ
DQ52
DST4
DFL16.1
Eμ
114,684,105
DFL16.1
VH7183.3
3'
5'
DQ52
Scale
chr12:
100 kb
mm9
114,650,000
114,700,000
114,750,000
114,800,000
114,850,000
114,900,000
B-cell (CD43-) 8w C57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-

Figure 22. DHS Tracks Demonstrate Eμ Sensitivity for Correct Genotypes. Five tracks of DHS sites that cover the *IgH* locus are displayed in the UCSC Genome Browser in each figure above. In figure A the browser image is the complete *IgH* locus on chromosome 12: 114,500,000 – 117,500,000. A zoomed in 100 kb view that displays the area between the Eμ enhancer and the first V_H gene, V_H7183.3, is in figure B. In both figures, the first track is mouse spleen B Cells ($CD43^-$, $CD11b^-$) from the Stamatoyannopoulous Laboratory at the University of Washington (UW) (GEO dataset: GSM1003813). The next track is mouse Pro-B cells from the laboratory of Meinrad Busslinger at Vienna Biocenter (VB) (GEO dataset: GSM932968). The next three tracks are $Rag8^{-/-}$, followed by B6D345 and $Rag2^{-/-}$ $E\mu^{-/-}$ in the bottom track. Results are representative from two independent DNase I digestion experiments for 20 units of DNase I digestion. The location of the Eμ deletion at position 114,665,501 is shown by the red arrow in figure B. Also shown in figure B are DQ52 (chr12:114,668,722) and DFL16.1 (chr12:114,720,385). A unique peak was identified at position 114,686,105, which is located between DQ52 and the first DSP gene segment and is reduced in $Rag2^{-/-}$ $E\mu^{-/-}$.

Figure 23

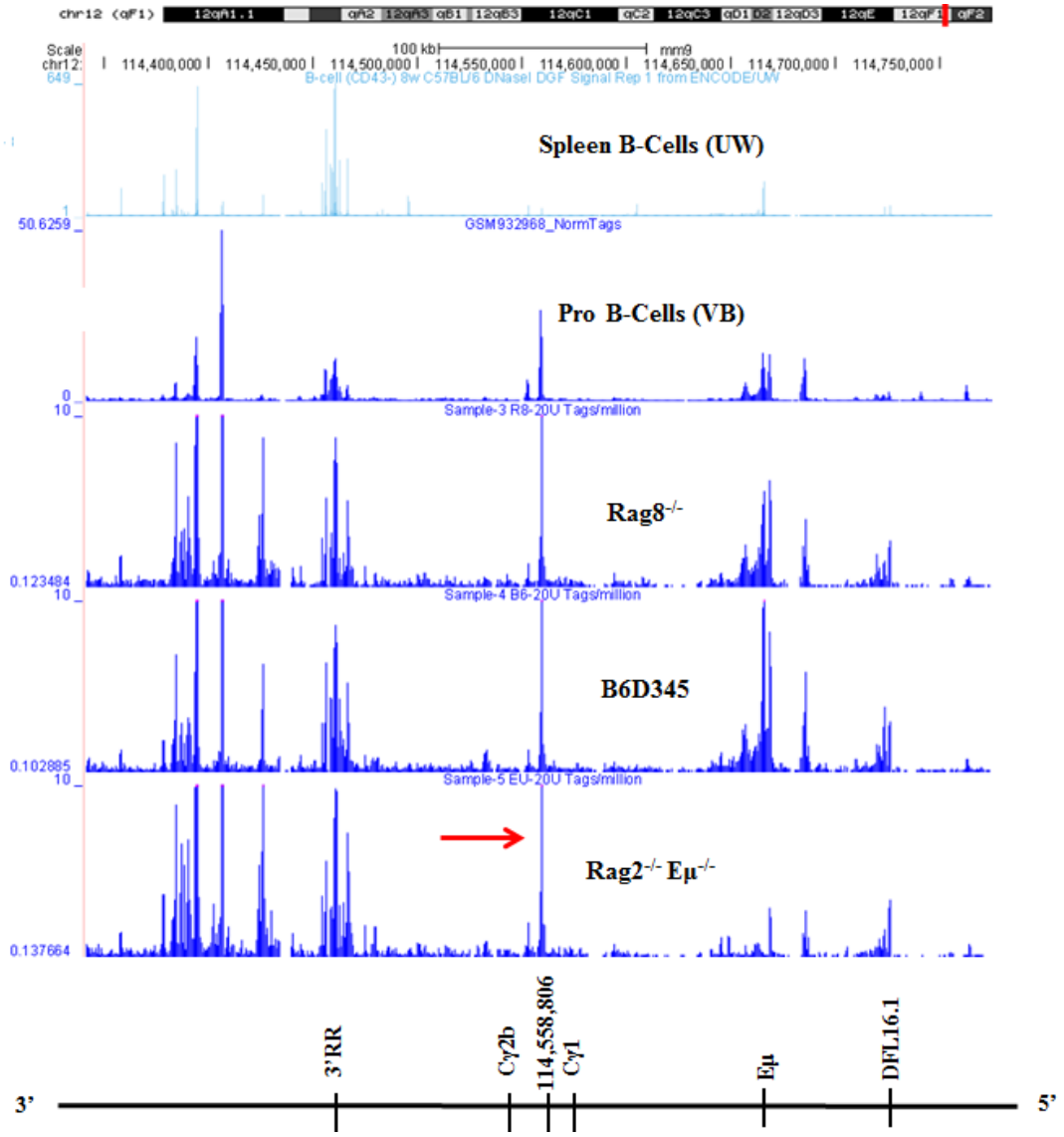

chr12 (qF1)
Scale
chr12:
100 kb
mm9
114,400,000
114,450,000
114,500,000
114,550,000
114,600,000
114,650,000
114,700,000
114,750,000
B-cell (CD43-) 8w C57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
649
1
50.6259
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
0
10
Sample-3 R8-20U Tags/million
Rag8-/-
0.123484
10
Sample-4 B6-20U Tags/million
B6D345
0.102885
10
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-
0.137664
3'RR
Cγ2b
114,558,806
Cγ1
Eμ
DFL16.1
3'
5'

Figure 23. Unique DHS peak identified between C-γ1 and C- γ2b. DHS peak location is between 114,558,806 - 114,560,261, as indicated by the red bar, on mouse chromosome 12 and is the largest peak within a 200 kb window in each of the three tracks at the bottom of the UCSC Genome Browser above. The first track is mouse spleen B Cells ($CD43^-$, $CD11b^-$) from the Stamatoyannopoulous Laboratory at the University of Washington (UW) (GEO dataset: GSM1003813). The next track is mouse Pro-B cells from the laboratory of Meinrad Busslinger at Vienna Biocenter (VB) (GEO dataset: GSM932968). The next three tracks are $Rag8^{-/-}$, followed by B6D345, and $Rag2^{-/-}$ $E\mu^{-/-}$ in the bottom track. Results are representative from two independent DNase I digestion experiments for 20 units of DNase I digestion.

Figure 24

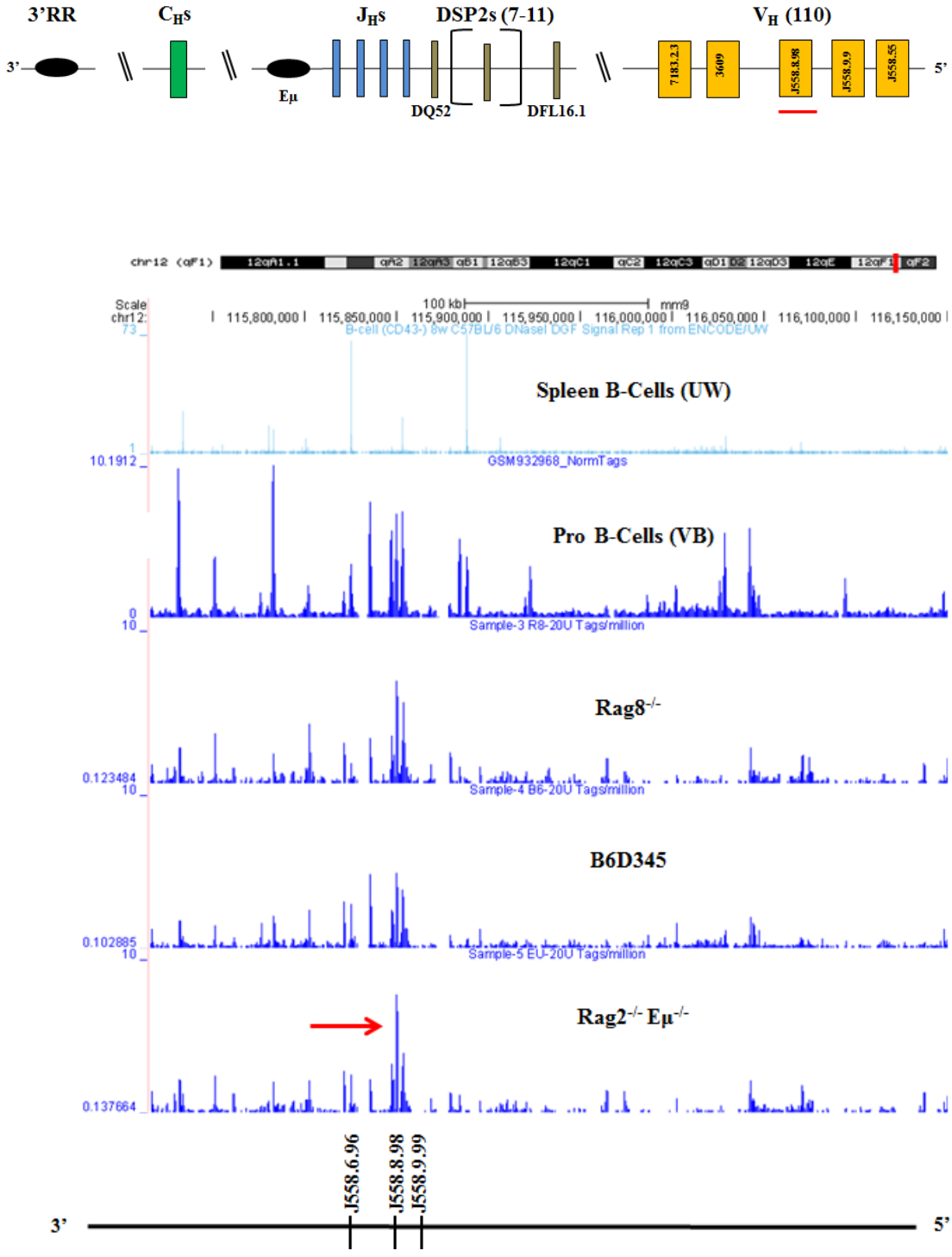

3'RR
C_HS
J_HS
DSP2s (7-11)
V_H (110)
3'
5'
Eμ
DQ52
DFL16.1
7183.2.3
3609
J558.8.98
J558.9.9
J558.55
Scale
chr12:
100 kb
mm9
115,800,000
115,850,000
115,900,000
115,950,000
116,000,000
116,050,000
116,100,000
116,150,000
B-cell (CD43-) 8w C57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-
J558.6.96
J558.8.98
J558.9.99
3'
5'

Figure 24. Unique DHS peak at J558.8.98. (A) Peak location is between 115,850,147 - 115,850,965 on mouse chromosome 12, as indicated by the red arrow. The first track is mouse spleen B Cells ($CD43^-$, $CD11b^-$) from the Stamatoyannopoulous Laboratory at the University of Washington (UW) (GEO dataset: GSM1003813). The next track is mouse Pro-B cells from the laboratory of Meinrad Busslinger at Vienna Biocenter (VB) (GEO dataset: GSM932968). The next three tracks are $Rag8^{-/-}$, followed by B6D345, and $Rag2^{-/-}E\mu^{-/-}$ in the bottom track. This is a significant peak that appears as the tallest in only the three tracks displayed below the UW and VB tracks in the UCSC Genome Browser above. Within the 100 kb window above, this peak is surrounded by a cluster of smaller peaks for the three bottom tracks. Results are representative from two independent DNase I digestion experiments for 20 units of DNase I digestion.

Figure 25

A

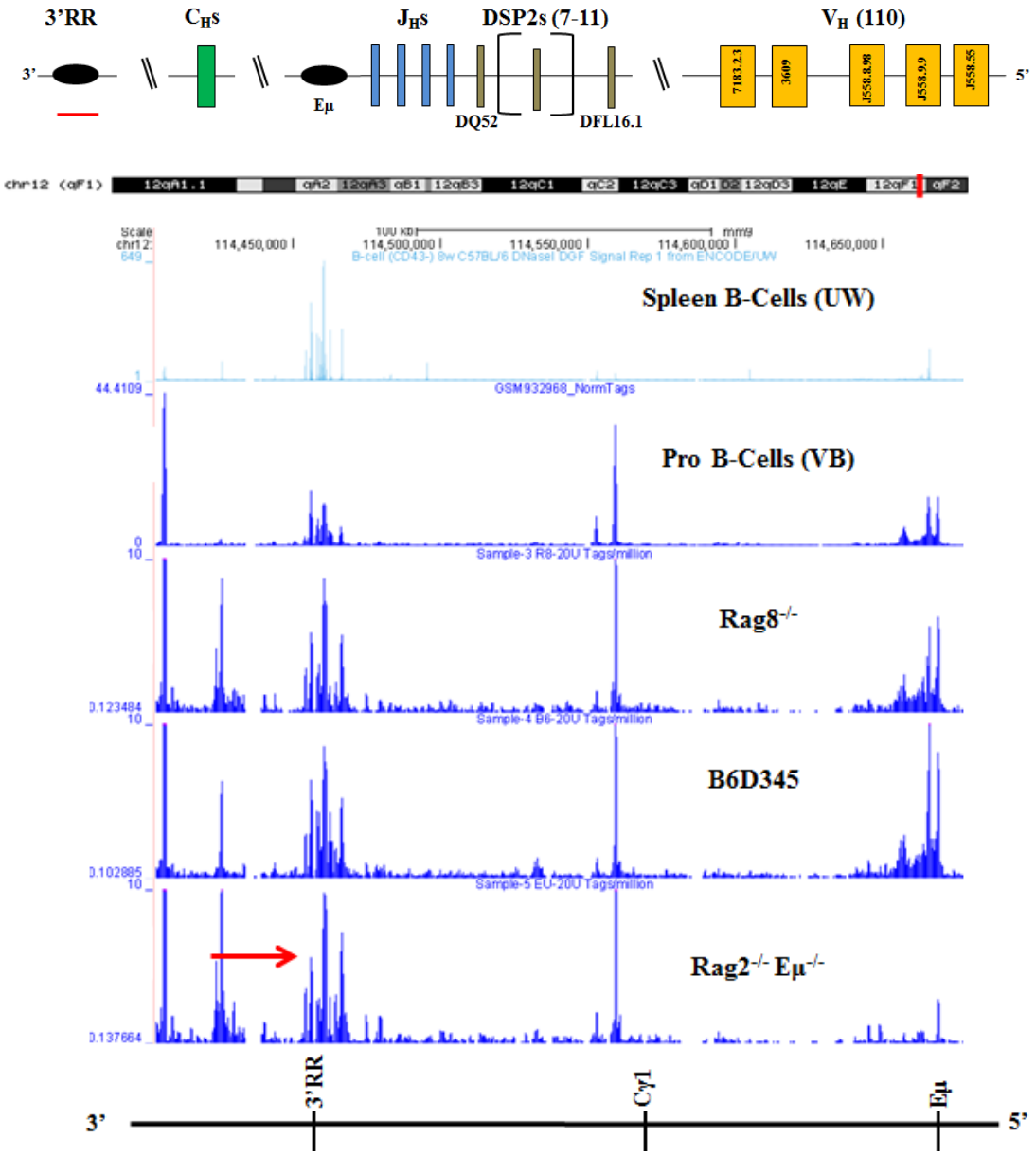

3'RR
C_Hs
J_Hs
DSP2s (7-11)
V_H (110)
3'
Eμ
DQ52
DFL16.1
7183.2.3
3609
J558.8.98
J558.9.9
J558.55
5'
chr12 (qF1)
12qA1.1
qA2
12qA3
qB1
12qB3
12qC1
qC2
12qC3
qD1
D2
12qD3
12qE
12qF1
qF2
Scale
chr12:
114,450,000
114,500,000
114,550,000
114,600,000
114,650,000
B-cell (CD43-) 8w C57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-
3'RR
Cγ1
Eμ
3'
5'

Figure 25

B

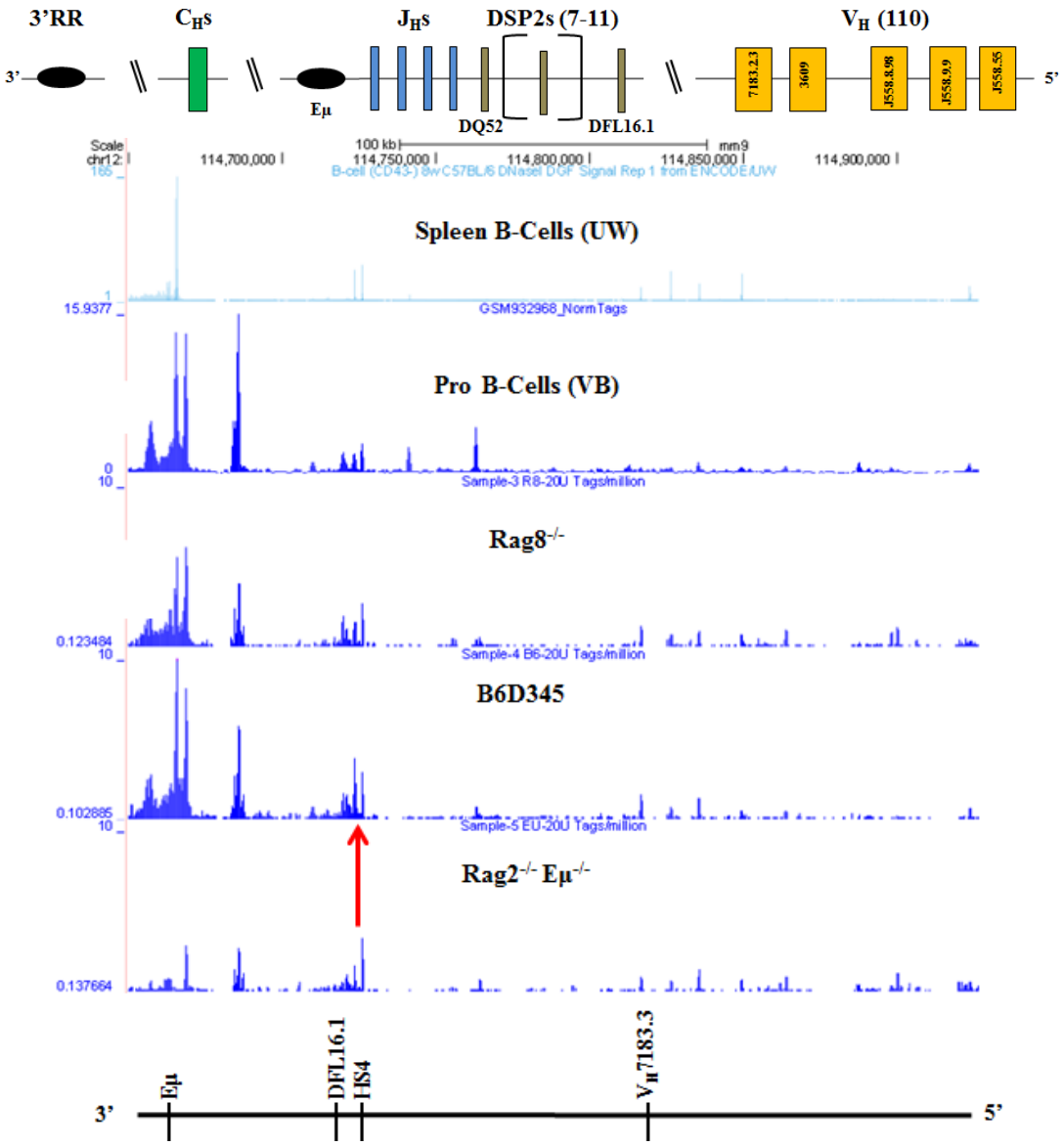

3'RR
C_Hs
J_Hs
DSP2s (7-11)
V_H (110)
3'
5'
Eμ
DQ52
DFL16.1
7183.2.3
3609
J558.8.98
J558.9.9
J558.55
Scale
chr12:
114,700,000
100 kb
114,750,000
114,800,000
114,850,000
mm9
114,900,000
B-cell (CD43-) 8wC57BL/6 DNaseI DGF Signal Rep 1 from ENCODE/UW
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-
Eμ
DFL16.1
HS4
V_H7183.3
3'
5'

Figure 25

C

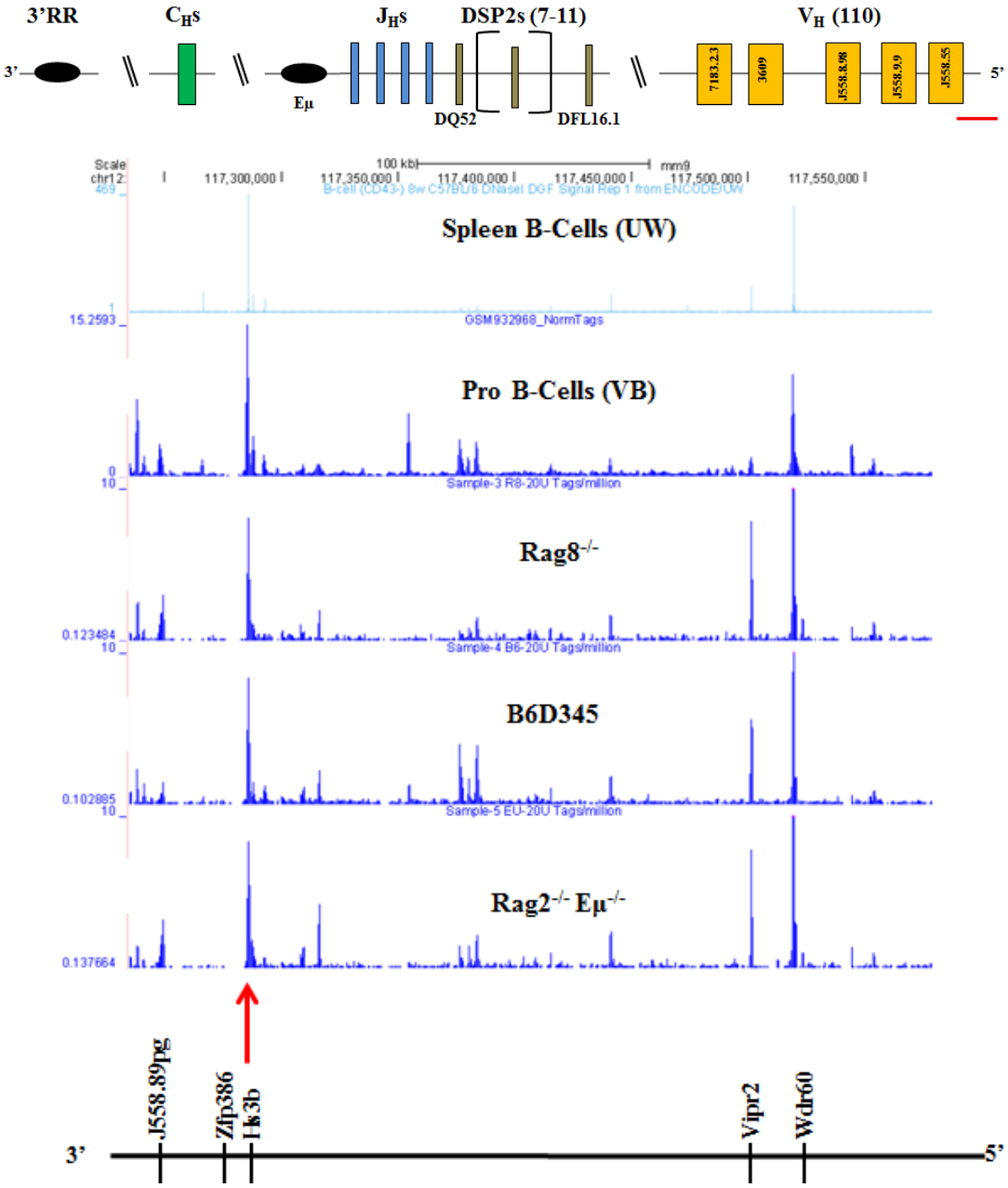

3'RR
C_Hs
J_Hs
DSP2s (7-11)
V_H (110)
3'
5'
Eμ
DQ52
DFL16.1
7183.2.3
3609
J558.8.98
J558.9.9
J558.55
Scale
chr12:
117,300,000
117,350,000
117,400,000
117,450,000
117,500,000
117,550,000
100 kb
mm9
Spleen B-Cells (UW)
GSM932968_NormTags
Pro B-Cells (VB)
Sample-3 R8-20U Tags/million
Rag8-/-
Sample-4 B6-20U Tags/million
B6D345
Sample-5 EU-20U Tags/million
Rag2-/- Eμ-/-
J558.89pg
Zfp386
Ik3b
Vipr2
Wdr60
3'
5'

Figure 25. DHS Peaks within the intergenic D_H to V_H region upstream of the last 5' V_H gene. UCSC genome browser display of DHS sites within and near the *IgH* locus. Above each figure is a schematic that shows the broad region of view within the *IgH* locus. Beneath each figure is a gene schematic of the specific region. (A) 100 kb view of the 3'RR to Eμ region of the *IgH* locus. (B) 100 kb view of the intergenic region between DFL16.1 and V_H 7183.2.3. The red arrow indicates the location of HS4as identified by Featherstone et al., 2010. (B) 30 kb region upstream of the last V_H gene, J558.89pg.195, near Zfp386. The red arrow is the location of Hs3b as identified by Pawlitzky et al., 2006.

Materials and Methods

The following methods and materials describe the protocols used to execute the experiments. Most experiments, including PCR, were carried out under sterile conditions in a tissue culture hood, and then transferred to a laboratory bench when appropriate. The exception is the Southern Blot procedure. Autoclaved tips were always used, including cut-tips to handle nuclei and other fragile samples.

Cell Culture Media

RPMI-RAG Media:

Bone marrow derived pro-B^+ cell lines were maintained in sterile tissue culture conditions. Media composition was as follows: RPMI-1640 medium (Invitrogen, catalog #11875-093) supplemented to contain 10% Fetal Bovine Serum, 1% Penicillin-Streptomycin-Neomycin (PSN) Antibiotic Mixture (Invitrogen, catalog # 15640-055), and 50 μM 2-mercaptoethanol. Media was filtered into a sterilized autoclaved bottle. Cell lines were maintained consistently at a density $0.5 \times 10^5 - 1.5 \times 10^6$ cells/ml. Prior to experiments, cells were >98% viable upon trypan blue stain. This media was designated "RPMI-RAG" media because of its components and use with cell lines associated with RAG deficiency.

OP-9 Media:

OP-9 media consisted of the following components: Minimum Essential Medium Alpha (Invitrogen, catalog number 11095-072), supplemented with 20% fetal bovine serum (Hyclone), 100 U/mL penicillin, 100 μg/mL streptomycin, and 1× L-glutamine

(Invitrogen); media was filtered into a sterilized autoclaved bottle and plated 1 or 2 days before use in 24-well plates (Corning) to achieve a final concentration of 2 x 10^5 cells/ml prior to transfer into a T-25 flasks. One day prior to primary pro-B cell plating, cells were plated in a T-25 flask at 3 x 10^5 cells/ml.

Primary Pro-B Cell Media:

Freshly isolated $CD19^+$ pro-B cells from mouse bone marrow tissue were plated on OP-9 cells, supplemented with 1ug IL-7, in media that contained the following: 500 ml OPTI-MEM (Invitrogen, catalog #11058-021), 4% Fetal Bovine Serum, and 1% Pen-Strep (Invitrogen). Media was filtered into an autoclaved bottle prior to use.

Cell Culture: Cell Lines, OP-9 and Primary Cells

Three mammalian bone marrow pro-B^+ derived cell lines were used in projects 1 and 2 of this thesis: B6D345, RAG2 deficient ($RAG8^{-/-}$), and $RAG2^{-/-}$ $E\mu^{-/-}$. Each of these cell lines were grown in the described "RPMI-RAG Media." The B6D345 pro-B cell line is from a C57BL6 mouse and has an inactive RAG1 allele (Ji et al., 2010), and was kindly provided by Dr. David Schatz (Yale University). Abelson transformed $RAG2^{-/-}$, referred to as $RAG8^{-/-}$ in this thesis, and $RAG2^{-/-}$ $E\mu^{-/-}$ were cultured in sterile tissue culture conditions and were maintained consistently at a density 0.5 x 10^5 – 1.5 x 10^6 cells/ml. Prior to experiments, cells were >98% viable upon trypan blue stain. OP-9 stromal cells, which are derived from bone marrow of the op/op mouse, support primary pro-B^+ cell growth in culture (Vieira and Cumano, 2004), and were a kind gift from Dr. Jagan Pongubala and were maintained in the described "OP-9 Media." These cells were not

grown longer than 10 days, unless used for primary-B cell experiments. Primary bone marrow derived $CD19^+$ pro-B cells were plated using a total of 3×10^6 cells on an OP-9 stromal layer plated 1 day prior at 3.0×10^5 cells/ml. Primary pro-B cell concentration during growth and expansion phase ranged from 1.5×10^6 to 2.0×10^6 cells/ml. Total number of primary pro-B cells never exceeded 40 million cells per T-75 Tissue Culture Flask (Corning, catalog # 430641). Media used for primary pro-B cell growth was the described "Primary Pro-B Cell Media." All cells were cultured in sterile tissue culture conditions, maintained in 37 ^{0}C incubators, and counted using a Hemocytometer.

Generation of Transgenic Mice

The BAC used to generate each transgenic mouse was derived from BAC RP23-458I14 (Invitrogen). The vector sequence is from the pBACe3.6, which is an 11 kb backbone that contains both a Sacb gene and PI-SceI linearization enzyme. The RP23-458I14 insert was estimated to be 181 kb based off PCR. BAC vectors were injected into C57BL\6 mice by pronuclear injection.

Mouse and Cell Line Genotyping

All the pro-B^+ cell lines and mouse colonies, including the transgenic mice, C57BL\6, J_HT, hs5,7-deficient, and Pax5+/- x $Rag2^{-/-}$. For the pro-B^+ cell lines, which includes B6D345, $Rag8^{-/-}$, and $Rag2^{-/-}E\mu^{-/-}$, the following primer sets were used: Eμ-5'(Up), Eμ-3' (Down), Emu5F, and Emu2R. A full outline of primers is given in the Figure 7 legend. Figure 7A below depicts primer locations in red bars, and key restriction sites with upside-down purple arrows. To genotype pro-B^+ cell lines, the following combinations

of primers were used: (a) Eμ-5' and Eμ-3'(b) Emu5F and Emu2R. Primer set (a) is outside of the XbaI-EcoR1 Eμ location and yields a PCR fragment size of 1,120 bp for $E\mu^{+/+}$ pro-B cells (B6D345 and $Rag8^{-/-}$), while this same primer set has a size of 900 bp in the Eμ deficient pro-B cell line, $Rag2^{-/-}E\mu^{-/-}$ (Figure 7B). The Eμ deletion is between the two Hinf1 enzyme sites in the $Rag2^{-/-}E\mu^{-/-}$ pro-B cell line. Using both primer sets allows detection and prevention of heterozygosity, mixed samples, or contamination.

For the transgenic mice, several different primer sets were used to test for the transgene in the J_HT background mice (Figure 7). These sets included the above listed primer sets, Eμ-5', Eμ-3', Emu5F, and Emu2R, and in addition, relocated primer sets, along with a specific primer that anneals to the Bac backbone (Sacb). Figures 7B-E are examples of genotyping the different transgenic mice with the different primer sets.

All PCR reactions for genotyping used the following reagents from Qiagen's HotStarTaq DNA Polymerase (Catalog # 203205) kit that is listed below. A PCR mastermix was set up using the following conditions:

10xPCR	5ul
MgCl2	2ul
Hot-Start Taq	0.5ul
dNTP	1ul
H20	39.5ul
Primer	2ul

All PCR Reactions, except J_HT, were executed under the following PCR cycling conditions:

Step	**Temperature**	**Length**
1	95 deg	15 min
2	94 deg	45 sec
3	58 deg	45 sec

4	72 deg	1:20 min
5	72 deg	10 min
6	4 deg	Forever

Steps #2-#5 were repeated 30 times.

J_HT PCR cycling conditions are as follows:

Step	Temperature	Length
1	94 deg	2 min
2	94 deg	30 sec
3	56 deg	1 min
4	72 deg	1:30 min
5	72 deg	10 min
6	4 deg	Forever

Steps #2-#5 were repeated 30 times.

PCR conditions to test D_H to J_H rearrangements are as follows:

Step	Temperature	Length
1	95 deg	15 min
2	94 deg	45 sec
3	62 deg	45 sec
4	72 deg	1:20 min
5	72 deg	10 min
6	4 deg	Forever

Steps #2-#5 were repeated 33 times.

Purification of CD19+ pro-B cells

Bone marrows of transgenic mouse strains were flushed using 19 G x 1in. (BD, catalog #305186) or 26 G x ½" needles (BD, catalog # 305111). ACK (Ammonium-Chloride-Potassium) Lysing Buffer was used for the lysis of red blood cells (Gibco, catalog # A10492-01). $CD19^+$ B cells were isolated by positive selection with anti-CD19 microbeads (Miltenyi Biotec). All selections were performed on a Miltenyi automacs separator according to the manufacturer's instructions. The purity of selected

populations was assessed by flow cytometry by staining in FACS buffer (PBS, 0.5% BSA, 0.01%NaN3, pH 7.2-7.4) containing the appropriate, pre-titered antibodies: CD19 PE-Cy7 (BD Biosciences, catalog # 561739), B220 FITC (BD Biosciences, catalog # 553087), A4.1 APC (Thy1.2) (Affymetrix, catalog # 17-0902-81), IgM-APC (Affymetrix, catalog # 17-5790-82), CD43 PE (BD Biosciences, catalog # 553271), CD19 FITC (BD Biosciences, catalog # 553785).

Real Time, Reverse Transcriptase PCR

Real-time PCR was performed on all DNase, ChIP, and transcription experiments using SYBR Green Universal Mix and the ABI Prism 7000 (Applied Biosystems). For ChIP, input DNA from purified chromatin before immunoprecipitation, along with immunoprecipitated (IP) DNA was first quantified using picogreen fluorescence kit (Molecular Probes, Eugene, OR).

All real-time PCR conditions were executed using the referenced amount of DNA at 50 ^{0}C for 2 minutes, 95 ^{0}C for 10 minutes, and then 40 cycles of 95 ^{0}C for 15 seconds and 60 ^{0}C for 1 minute. Data was collected at 60 ^{0}C. The endpoint used in real-time quantification is the PCR cycle number that crosses an arbitrarily placed signal threshold, and this threshold is a function of the amount of target DNA present in the starting material. For ChIP experiments, the threshold was set at a point at which the real-time PCR amplification was linear. The fold difference for any target sequence was calculated to measure the fold enrichment of the target sequence in the immunoprecipitate compared to the total sample (input) (Litt et al., 2001): $2^{-\Delta Ct}$($\Delta Ct = C_t[IP] - C_t[Input]$. For transcription analysis, one microgram of total RNA was reverse transcribed using

Superscript III (Invitrogen) and random hexamers. Real-time PCR was performed using 0.5 ul of cDNA for both +RT and –RT samples.

Southern Blot

Mouse DNA was isolated from tissues as outlined previously (Strauss, 2001). Ten micrograms of genomic DNA was completely digested overnight at 37 °C with different restriction enzymes (5 units) in a total volume of 500ul. Samples were then precipitated the next day by ethanol precipitation and resuspended in 15 µl 10mM Tris pH 7.5. Subsequently, samples underwent electrophoresis in 1x TBE at 35V for 16 hours. The gel was then treated with 0.25 M HCl to fragment the high molecular weight DNA into small pieces. After this, the DNA was denatured and transferred overnight onto a GeneScreen Plus (PerkinElmer) membrane using alkaline transfer buffer (0.4 M NaOH, 1 mM EDTA). To probe for the DNA on the blot, genomic DNA was amplified using primers for the area of interest, and the DNA was radiolabeled with [32P]–α-deoxy-CTP using the NEblot kit (New England Biolabs). Blots were hybridized overnight at 62°C, washed the next, and the band patterns were detected using phosphorimaging screens and a Storm 860 phosphorimager (GE Healthcare Life Sciences).

Chromatin Immunoprecipitation

ChIP experiments on modified histones and RAG-1/-2 were performed using methods similar to those described previously (Ji et al., 2010; Chakraborty et al., 2009; Chowdhury and Sen, 2001; Lu et al., 2004). Approximately 10 million $CD19^+$ pro-B bone marrow cells, maintained prior at a cell concentration of 1 x 10^6 cells/ml, were

cross-linked for 5 minutes using 1% formaldehyde and quenched with 0.125 M glycine on a steady rocker. The cells were then lysed with SDS buffer (1% SDS, 50 mM Tris-HCl pH 8, 10 mM EDTA) and sonicated twice on ice/water using a Bioruptor to reduce the DNA length to approximately 500 – 1000bp. The debris was removed by centrifugation and the sonicated chromatin was then placed in ChIP dilution buffer (167 mM NaCl, 1.2 mM EDTA, 1.1% TX-100). Sonicates were pre-cleared with 5ug of nonspecific IgG, an input sample was saved, and immunoprecipitated with the following antibodies: anti-H3K9ac (Millipore, 06-942), anti-H3K9me2 (Millipore 07-441) anti-H3K4me3 (Active Motif, 39159) and rabbit immunoglobulin G (IgG; Santa Cruz, sc-2027). Immune complexes were then collected with protein A, washed, and eluted with sodium bicarbonate-SDS (1% SDS containing 0.1 M $NaHCO_3$), and the cross-links were reversed by heating at 65 °C for 6 hours, followed by proteinase K treatment, two phenol extractions, and ethanol precipitation. Chromatin, including the input and immunoprecipitated DNA samples, was quantified using fluorometrically using PicoGreen (Molecular Probes-Invitrogen, P7589). Real-time PCR was used to determine the abundance of target sequences in the immunoprecipitated relative to the input according to manufacturer's instructions on an ABI 7500 apparatus (Applied Biosytems). Each real-time PCR used 200 pg DNA, performed in duplicates, and each ChIP was done in duplicates or triplicates. The oligonucleotide sequences are presented in the Materials and Methods section under Primers and Probes. The abundance of RAG-1 or RAG-2 at specific genomic loci ($IP/Input_{corr}$) was calculated as previously described (Ji et al., 2010).

Genome-wide DNase I Hypersensitivity

DNase Hypersensitivity was performed using methods adapted from those described previously (John et al., 2013). Modifications to the conditions of previously described methods were made specifically for primary B-cells, B-cell lines, and consistent results. Buffers were always prepared prior to experiment. Concentration of cells in culture for experiment was always between 1.0 – 1.5 x 10^6 cells/ml. The source of the DNase I enzyme was Worthington-Biochem (#DPFF), and powder was dissolved in sterile nuclease-free water to a concentration of 1 unit/ul, aliquoted in 20ul volume and placed in -80^{0}C. Experiments were always performed using a limiting range of DNase I concentrations to liberate the most accessible regions in chromatin, always including undigested "0 units" and over-digested >40 units. A total of 10 million cells were used for one DNase I digestion and titration, and the enzyme was diluted and titrated prior to application to cells. Experiment efficiency was greatest when the maximum number of cells used for one experiment did not exceed 50 million.

Briefly, cells were pelleted and washed twice with cold dulbecco's phosphate-buffered saline that contained EDTA-free protease inhibitor (Roche 11873580001). Cells were then suspended in cold 1x DHS-B buffer (100mM Tris-HCl (pH 7.4), 100mM NaCl, 30mM CaCl2, 0.5 mM spermidine, protease inhibitor) in a total of 2ml and stored on ice. Nuclei was then purified and released by hypotonic lysis in the presence of 0.5% NP-40. After this point, only sterile, cut tips were used to handle the nuclei. The nuclei was then washed once with 1x DHS-B buffer. During this wash, titration of DNase I enzyme took place. It was essential to minimize the time of thawing the DNase I enzyme and applying to the reaction tubes. Nuclei were then diluted and digested with different

amounts of DNase 1 (Worthington) for 3 min at 37 °C in a total volume of 250ul of 1x DHS-B buffer. Reaction was terminated with an equal volume of stop buffer (0.5M EDTA, (pH 8.0 (10mM final)), 1x DHS-B buffer, 20% SDS (2% Final), Proteinase K (25 µg/ml final concentration)), and incubated at 55 °C for a minimum of 2 hours. Next, RNase A (Invitrogen 20 mg/ml RNase A) was added and sample was incubated at 37 °C. After DNase I treatments, the samples are subjected to careful phenol-chloroform extractions using 1.5 ml phase lock gel heavy tubes (5 Prime # 2302810) and stored at +4 °C. At this point, a small aliquot of DNA was recovered and ethanol precipitations were performed for gel electrophoresis and quantitative PCR (qPCR). A good DNase I experiment was justified by both a gel that showed a parent band with minimal, yet increasing digestion in most samples, and a qPCR that demonstrated the shift in CT values described below. Control (untreated) have no digestion characteristics as observed in both gel electrophoresis and qPCR. Significant smearing below the parent DNA band demonstrates over digested chromatin, endogenous nuclease activity, and/or sick cells. Lack of a parent band is indicative of either poor buffer prep or inaccurate processing of samples during the digestion stage. Control samples were processed as outlined above but without any DNase I treatment.

DNase I accessibility was determined by qPCR of the DNA samples (McArthur et al., 2001), with several modifications. DNase-treated DNA was quantified fluorometrically using PicoGreen (Molecular Probes-Invitrogen, P7589) and then diluted. Next, 200 pg was subjected, in duplicates, to real-time PCR using primers of similar amplicons size that cover either a hypersensitive or insensitive gene regions. PCRs were performed using SYBR Green PCR Master Mix (Applied Biosystems #4368702)

according to manufacturer's instructions on an ABI 7500 apparatus (Applied Biosystems). DNase I sensitivity was measured by the percent DNA remaining compared to the untreated sample. DNA content was normalized to a known DNase I-resistant locus (Nfm). As the number of DNase I units increases, the DNase I hypsensitive peaks are enriched compared to the closed, intergenic region. The goal for sequencing is to select an optimal DNase I concentration for the treated sample where 1% of the genome has been cut and the hypersensitive sample shows selective degradation and the CT value increases, but there is little or no change in the CT value for the primer that flanks the insensitive region. Experiments were duplicated to confirm quality of DNase I enzyme and reproducibility.

Once the optimum concentration of DNase I is determined, the treated fragments are purified by size selection over a sucrose gradient to prepare the correct size fragments for sequencing. To accomplish this, the DNase I treated nuclei are separated from the bulk of genomic DNA by a 24 hour x 25,000 rpm sucrose gradient ultracentrifugation (9% Sucrose, 5M NaCl, 1M Tris-HCl (pH 8.0), 10ml 0.5M EDTA), as previously described (John et al., 2013). Samples are then carefully collected in 500ul aliquots and the DNA isolated by ethanol precipitation. The size of these fragments is then completed by gel electrophoresis, and those fractions that are less than 500 bp are pooled together to construct the sequencing libraries for the Illumina Genome Analyzer. The DNA is ligated with adaptors and amplified to generate a library of fragments for high-throughput sequencing. A size selection step is included to limit the sequencing to fragments of 200-400 bp, which improves the signal to noise on the sample to be sequenced. Once the

DHS libraries have been sequenced and processed, they can be visualized using the UCSC Genome Browser (University of California Santa Cruz).

Probes and Primers

Genotyping

Eμ trio 3' in	ctcagcctggactttcggtt
Eμ trio 3' out	accaccaaccagcatgttca
Eμ trio 5' out	aggaatgggagtgaggctct
Arms_FP	ggggtgtatcctctggctacaggag
Arms_RP	tgcagccatgaataacaaagag
Eusense2	gagtgaggctctctcataac
Eμ FP	ggaatgggagtgaggctctctc
Eμ RP	ctgcaggtgttctggttctgatcgg
Eμ-3' (Down)	gtgaagccgttttgaccaga
Eμ-5'(Up)	ttcaggaccacctctgtgac

Southern Blot

5primejht_for.6	gaaaggccaaagcttcctct
5primejht_for.7	gcaagccaaataagcagcatt
5primejht_for.8	tttttcctggcacatgtccata
5primejht_rev.6	aggaaaggtggtgtctgtgg
5primejht_rev.7	accttccgagcccacagtct
5primejht_rev.8	gcagtaggtgcttagggagcat
bac_cmu_for.1	aagggcttctaagccagtcc
bac_cmu_for.2	ccagtccacatgctctgtgt
bac_cmu_for.4	tttcaagaccactttctgagtattcatt
bac_cmu_rev.1	gagttagtccagccgacctg
bac_cmu_rev.2	aatcccaggattgccttctt
bac_cmu_rev.4	ttcctcaaagtaccatgagctctatg
BacEnd_For1	ccccaccaaactgagacact
BacEnd_For2	ccctattgatgacgagttgg
BacEnd_Rev1	gactccccgtccacagtaaa
BacEnd_Rev2	aaatgcaaagcaaacctcag

DNase Hypersensitivity

B2M1F	ggccaggggtttaacttctc

B2M1R	cttggggtttctgcttatcg
B2M2F	ccaaaacccatggacagaac
B2M2R	ccttgcttctagcaggtttctc
B2M3F	tgctcttaattgtcctggcttt
B2M3R	ggaccgccttcatttattca
B2M4F	tgggaaagtccctttgtaacc
B2M4R	gcgcgcgctcttatatagtt
B2M5F	tggtgcttgtctcactgacc
B2M5R	cgggaaggagaactaaggaga
B2M6F	tgaggctacagggtggatg
B2M6R	ttttgcgctcagggagtcta
B2M7F	ctctgcgatgtttccaaagc
B2M7R	ctacctctgcctcccaagtg
B-globin HS1_FOR	agattatattgccatggtacacttgaa
B-globin HS1_REV	actggaccaattttctccctcc
B-globin HS2_FOR	agtgtcagcatattaccgatgttcc
B-globin HS2_REV	cacacagcaaggcagggtc
B-globin HS3_FOR	tgtaagtgtaaattttggagcacagg
B-globin HS3_REV	ctgaaagactaaagttcccggc
B-globin HS4_FOR	tgtttgtggtttttctgttgtatgttt
B-globin HS4_REV	aagagcagaaaggaattaaatacacaca
B-globin HS6_FOR	cagagcattgttgaaagatgagga
B-globin HS6_REV	gggttagcagaaatgtagagctcc
B-globin_DHStrial_For	gggggttaattttaaagacctaca
B-globin_DHStrial_Rev	acaggactccagcagatcgt
Chr12_DHSHi_114.39For	atgagagggaggatggcttt
Chr12_DHSHi_114.39Rev	aatcacagttccggatgagg
Chr12_DHSHi_117For	acaccaggcacagaccattc
Chr12_DHSHi_117Rev	ggcaagtccacactcagagc
Chr12_DHSHi_EmuFor	tggggaaactagaactactcaagc
Chr12_DHSHi_EmuRev	ttaaccgaggaatgggagtg

ChIP / Transcription

DFL(+1.2)FP	cctgcaaacagagcttacca
DFL(+1.2)RP	tccagtactgctgcttggtc
DFL(-0.1)FP	gatagaagcacagaagtgaataatc
DFL(-0.1)RP	ggttttgcaggtgtagagagcattg
DFL(-6.0)FP	tagttttcaaatgttatccccgg
DFL(-6.0)RP	gtgactctagtgcatatgtaaa

DFL16.1 (+3)_FP	agcctaaaagccactcacga
DFL16.1 (+3)_RP	tcttggaacactggggaaac
DFL16.1FP	acacctgcaaaaccagagaccata
DFL16.1RP	gatctgactttagggtggttgga
dp2xanti1	catttggtcctatgtgttct
DP2Xsense2	acaaccctactgagaaccag
dp2xsense2	acaaccctactgagaaccag
DQ(-0.1)FP	gcgactgttttgagagaaatcattgg
DQ(-0.1)RP	agggactctgaggcttaaacc
DQ(-1.0)FP	gagcagaatcccagccaagag
DQ(-1.0)RP	gaggtctagaggctcccttcc
DQ(-2.0)FP	gcctgccattggatcccct
DQ(-2.0)RP	ggtatcattctgcctagggacatc
DQ(-3.3)FP	atgtctgtgggtatgctcctgac
DQ(-3.3)RP	tggccctgttgcctaaggatg
DQ52_FP	ccctgtggtctctgactggtg
DQ52_RP	gatttctcaagcctctctacttcctc
dsp 2.11-1.5 rp	tctgcctttcagcccctcttct
DSP2.2 FP	ccaggcaatgttcctgcagaatctc
DSP2.2 RP	ctgggtgagagactattgggatctgtt
DSP2.2(-1.0)FP	tgtccagtagaaaggtgcacatg
DSP2.2(-1.0)RP	atatttgggtttcagtgctgatgacc
DSP2.2(-1.4)FP	tccagtccacactcatcaccttggc
DSP2.2(-1.4)RP	gttgctctcctttagcccaagttctg
dsp2.2(2)-2.4 f	gtcttacccacccctgcccccaaga
DSP2.2(-3.0)FP	tcatgtaaaagtggtaaacacctg
DSP2.2(-3.0)RP	ctgatgctgttaaacacctccta
DSP2.3Fp1	ctgtcagcgccatcttgtaa
DSP2.3Fp2	gcaatgcccacagttacctt
DSP2.3Fp3	ccttccctgagtctctgtgc
DSP2.3Fp4	gtctcctttccctcctcacc
DSP2.3Fp5	cccacagttaccttccctga
DSP2.3Rp1	aaacagcctcccaatgtcac
DSP2.3Rp2	taggcagagtccccttctga
DSP2.3Rp3	taggcagagtccccttctga
DSP2.3Rp4	gcctcagtcctgatgctagg
DSP2.3Rp5	taggcagagtccccttctga
DSP2.9 FP	ccaccagaccatgttcctgcat
DSP2.9 RP	gtacctgatctatgtcaatatctgtcccata

DSP2.9(-5.5)FP	gagtaaagaagaattgtaagggtagactgtc
DSP2.9(-5.5)RP	gagaaggaatcctgttgaacatcgc
DSP2s FP	tgttaccttacttggcagggattt
DSP2s RP	tgggtttttgttgctggatatatc
Emu+1_FP	aagggcttctaagccagtcc
Emu+1_RP	gcagaagccacaaccataca
Emu1F	ggctgagagaagttgggaaa
Emu1R	caaatgagcctccaaagtcc
Emu2F	atccgatgcatagggacaaa
Emu2R	tttgctcagcctggactttc
Emu2R	ttgagaccgaggctagatgc
Emu3F	cctgcaaaagtccagctttc
Emu3R	cttgaaaaccctctcacatttt
Emu4F	tgtccaaaattcttgtcaatcag
Emu4R	ttgttaaaagtcagttctgaataggtt
Emu5F	ggtcatgtggcaaggctatt
Emu5F	atccgatgcatagggacaaa
Emu5R	tttgctcagcctggactttc
Emu6F	acaccacctgggtaatttgc
Emu6R	tggggaaactagaactactcaagc
Emu7F	ggttatgtaagaaattgaaggacttt
Emu7R	aatgcttcgggtattggaaa
Emu8F	cagttgaacatgctggttgg
Emu8R	tggcccattcaacaataagc
Eusense2	gagtgaggctctctcataac
Eμ FP	ggaatgggagtgaggctctctc
Eμ RP	ctgcaggtgttctggttctgatcgg
Eμ trio 3' in	ctcagcctggactttcggtt
Eμ trio 3' out	accaccaaccagcatgttca
Eμ trio 5' out	aggaatgggagtgaggctct
Eμ-3’ (Down)	gtgaagccgttttgaccaga
Eμ-5’(Up)	ttcaggaccacctctgtgac
Fk1f	ggcttttgaggcaacactat
Fk1r	cattggcagaaagctctcataca
Fk2f	ggattttactatataactatgctatca
Fk2r	agaaagtaagggatcacgtgtaatacaac
Fk3f	gtaccatgtgtttgtgtgaagtaga
Fk3r	tttttgaggatttccatcagcat
hs6_for	gatacacgcgcagctgctggcttgc

hs6_rev	ggacacaggttcgaggtgctcag
ice_bac_cmu_for.3	gctatgctacgctgtgttgg
ice_bac_cmu_for.5	ggttggtggttgagaggacact
ice_bac_cmu_rev.3	gccagccagctctactcact
ice_bac_cmu_rev.5	attggttaacaggcaacatttttct
IMR1746	gaggttccgctacgactctg
IMR3104	ccggacaagtttttcatcgt
IMR8162	tggatgtggaatgtgtgcgag
JH1FP_C	gctcaggtcactcaggtcagg
JH1RP	ggcctgacatggggagatctgagaatatc
JH1RP_C	gagaatatcttttcccgtttcaga
JH2FP_C	tactttgactactggggc
JH2RP	gctgaatagaagagagaggttgtaaggac
JH2RP_C	ccctagtccttcatgacc
Jh4 FP_C	caccaggaattggcataa
Jh4 RP_C	cctgaggagacggtgact
JH4FP	caccaggaattggcataa
JH4RP	cctgaggagacggtgact
J_ht_seq_for	cctgataggcacccaagtac
J_ht_seq_rev	ttagaaattaaagacactaaagtc
J_ht_seq_rev_wt	atccacccttctgatgcttg
Nfm1F	gactggcaagcctagaatgc
Nfm1R	aagctggcttcagtccaaaa
Nfm2F	cgctctcgagtgtcgtgtc
Nfm2R	acgacctcagcagctaccag
Nfm3F	gccgtggagatgtctgtctt
Nfm3R	ggagcaatcacgaagaggag
Nfm451_For	gctgggtgatgcttacgacc
Nfm451_Rev	gcggcatttgaaccactctt
Nfm4F	cttgagccttctcgtggttc
Nfm4R	gaagcagagatccaggcact
Nfm5F	gacgaggactggctgaagtc
Nfm5R	cgtgtcctcctcctacaagc
RagA	gggaggacactcacttgccagta
RagB	agtcaggagtctccatctcactga
RagNeo	cggccggagaacctgcgtgcaa
SacB anti	atgttcattaagaagcttggcgc
SacB sense	gccacagttctgacttttacgac
Sacb_For	gccacagttctgacttttacgac

Sacb_Rev	atgttcattaagaagcttggcgc
vd_dn_anti1	taacactacagcacatgtgt
vd_dn_anti2	cagcttctacatgcctaccc
vd_up	tcacaatgcggttcggatat
VDUPgS2	tcccaatacctcttataacc
y-actin FP_C	gacacccaaccccgtgacg
y-actin RP_C	gcggccatcacatcccag
β-globin FP	gccttgcctgttcctgctc
β-globin RP	attgagccctttactctctctgttc
γ-actin FP	gacacccaaccccgtgacg
γ-actin RP	gcggccatcacatcccag

References:

Abarrategui I., Krangel M.S. (2006). Regulation of T cell receptor-a gene recombination by transcription. Nat Immunol., 7:1109–15.

Afshar, R., Pierce, S., Bolland, D. J., Corcoran, A., Oltz, E. M. (2006). Regulation of IgH gene assembly: Role of the intronic enhancer and 5'∆y52 region in targeting DHJH recombination. J. Immunol., 176(4), 2439–2447.

Alessandrini, A., Desiderio, S. V. (1991). Coordination of immunoglobulin DJH transcription and D-to-JH rearrangement by promoter-enhancer approximation. Mol. Cell. Biol., 11(4) 2096–2107.

Alt F. W., Rosenberg N. Casanova R. J., Thomas E., Baltimore D. (1982). Immunoglobulin heavy-chain expression and class switching in a murine leukaemia cell line. Nature, 296: 325–331.

Alt F.W., Yancopoulos G.D., Blackwell T.K., Wood C., Thomas E., Boss M., Coffman R., Rosenberg N., Tonegawa S., Baltimore D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J., 3:1209-1219.

Antoch M.P., Song E.J., Chang A.M., Vitaterna M.H., Zhao Y., Wilsbacher L.D., Sangoram A.M., King D.P., Pinto L.H., Takahashi J.S. (1997). Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. Cell, 89(4):655-67.

Atkinson M.J., Chang Y., Celler J.W., Huang C., Paige C.J., Wu G.E. (1994). Overusage of mouse DH gene segment, DFL16.1, is strain-dependent and determined by cis-acting elements. Dev Immunol., 3(4):283-95.

Banerji J., Olson L., Schaffner W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell, 33(3):729-40.

Banerji, J., Rusconi, S., Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell, 27, 299–308.

Bassing C.H., Swat W., Alt F.W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. Cell, 109(Suppl):S45–55.

Bates J. G., Cado D., Nolla H., Schlissel M. S. (2007). Chromosomal position of a VH gene segment determines its activation and inactivation as a substrate for V(D)J recombination. J. Exp. Med., 204, 3247–3256.

Birshtein B.K. (2014). Epigenetic Regulation of Individual Modules of the immunoglobulin heavy chain locus 3' Regulatory Region. Front Immunol., 21;5:163.

Boboila C., Alt F.W., Schwer B. (2012) Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. Adv Immunol. 116:1–49.

Bolland D.J., Wood A.L., Afshar R., Featherstone K., Oltz E.M., Corcoran A.E. (2007). Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. Mol Cell Biol. 27:5523–33.

Bolland D.J., Wood A.L., Corcoran A.E. (2009). Large-scale chromatin remodeling at the immunoglobulin heavy chain locus: a paradigm for multigene regulation. Adv Exp Med Biol. 650:59-72.

Bolland D.J., Wood A.L., Johnston C.M., Bunting S.F., Morgan G., Chakalova L., Fraser P.J., Corcoran A.E. (2004). Antisense intergenic transcription in V(D)J recombination. Nat Immunol 5:630–637.

Bondarenko V.A., Liu Y.V., Jiang Y.I., Studitsky V.M. (2003). Communication over a large distance: enhancers and insulators. Biochem. Cell Biol. 81:241–251.

Bories J.C., Demengeot J., Davidson L., Alt F.W. (1996). Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility. Proc. Natl. Acad. Sci. USA. 93:7871–7876.

Bouvier G., Watrin F., Naspetti M., Verthuy C., Naguet P., Ferrier P. (1996). Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. Proc. Natl. Acad. Sci. USA. 93:7877–7881.

Boyle A.P., Davis S., Shulha H.P., Meltzer P., Margulies E.H., Weng Z., Furey T.S., Crawford G.E. (2008). High resolution mapping and characterization of open chromatin across the genome. Cell. 132:311–322.

Bulger M, Groudine M. (2010). Enhancers: The abundance and function of regulatory sequences beyond promoters. Dev Biol. 339: 250–257

Burnet, F.M. (1957). A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. Aust J Sci. 20(3): 67-69.

Chakalova L., Debrand E., Mitchell J.A., Osborne C.S., Fraser P. (2005). Replication and transcription: shaping the landscape of the genome. Nat Rev Genet. 6(9):669-77.

Chakraborty T, Chowdhury D, Keyes A, Jani A, Subrahmanyam R, Ivanova I, Sen, R. (2007). Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. Mol Cell. 27:842–50.

Chakraborty T., Perlot T., Subrahmanyam R., Jani A., Goff P.H., Zhang Y., Ivanova, I., Alt, F.W., Sen, R. (2009). A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. J Exp Med, 206:1019–27.

Chang Y., Paige C.J.,Wu G.E. (1992). Enumeration and characterization of DJH structures in mouse fetal liver. EMBO J. 11:1891–9.

Chatterjee S., Lufkin T. (2012). Regulatory genomics: Insights from the zebrafish. Curr Top Genet. 5:1-10.

Chatterjee S., Ju Z., Hassan R., Volpi S.A., Emelyanov A.V., Birshtein B.K. (2011). Dynamic changes in binding of immunoglobulin heavy chain 3' regulatory region to protein factors during class switching. J Biol Chem. 286:29303–12.

Chen J., Trounstine M., Alt F.W., Young F., Kurahara C., Loring J.F., Huszar D. (1993). Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the JH locus. Int. Immunol. 5:647–656.

Chen T.A., Allfrey V.G. (1987). Rapid and reversible changes in the nucleosome structure accompany the activation, repression and superinduction of the murine proto-oncogenes c-fos and c-myc. Proc. Natl Acad. Sci. USA, 84, pp. 5252–5256.

Chen T.A., Sterner R., Cozzolino A., Allfrey V.G. (1990). Reversible and irreversible changes in nucleosomal structure along the c-fos and c-myc oncogenes following inhibition of transcription. J. Mol. Biol., 212, pp. 481–493.

Cho K.W. (2012). Enhancers. Wiley Interdiscip Rev Dev Biol. Jul-Aug;1(4):469-78.

Chowdhury M., Forouhi O., Dayal S., McCloskey N., Gould H.J., Felsenfeld G., Fear D.J. (2008). Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. Proc Natl Acad Sci U S A. 105(41):15872-7.

Chowdhury, D., Sen, R. (2001). Stepwise activation of the immunoglobulin mu heavy chain gene locus. EMBO J. 20, 6394–6403.

Chowdhury, D.,Sen R. (2004). Regulation of immunoglobulin heavy chain gene rearrangements. Immunol Rev 200:182-196.

Clark, M.R., Cooper, A.B., Wang, L.D., Aifantis, I. (2005). The pre-B cell receptor in B cell development: recent advances, persistent questions and conserved mechanisms. Curr Top Microbiol Immuno 290: 87-103.

Cohn, M., Mitchison, N.A., Paul, W.E., Silverstein, A.M., Talmage, D.W., Weigert, M. (2007). Reflections on the clonal-selection theory. Nat Rev Immunol 7(10): 823-830.

Corcoran A.E., Riddell A., Krooshoop D., Venkitaraman A.R. (1998). Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. Nature 391:904–907.

Crawford G.E., Holt I.E., Whittle J., Webb B.D., Tai D., Davis S., Margulies E.H., Chen Y., Bernat J.A., Ginsburg D., Zhou D., Luo S., Vasicek T.J., Daly M.J., Wolfsberg T.G., Collins F.S. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. 16:123–131.

Curry J.D., Geier J.K., Schlissel M.S. (2005). Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. Nat Immunol 6:1272–9.

Daniel J.A., Santos M.A., Wang Z., Zang C., Schwab K.R., Jankovic M., Filsuf D., Chen H.T., Gazumyan A., Yamane A., Cho Y.W., Sun H.W., Ge K., Peng W., Nussenzweig M.C., Casellas R., Dressler G.R., Zhao K., Nussenzweig A. (2010). PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. Science. 329(5994):917-23.

Dillon, N. (2004). Heterochromatin structure and function. Biol. Cell. 96, 631–637.

Dudley, D. D., Chaudhuri J., Bassing C. H., Alt F. W. (2005). Mechanism and Control of V(D)J Recombination versus Class Switch Recombination: Similarities and Differences. Adv Immunol 86:43-112.

Duncan I.W. (2002). Transvection effects in Drosophila. Annu Rev Genet. 6:521-56.

Ebert A., McManus S., Tagoh H., Medvedovic J., Salvagiotto G., Novatchkova M., Tamir I., Sommer A., Jaritz M., Busslinger M. (2011). The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. Immunity 34:175–87.

Ebert A., Medvedovic J., Tagoh H., Schwickert T.A., Busslinger M. (2013). Control of antigen receptor diversity through spatial regulation of V(D)J recombination. Cold Spring Harb Symp Quant Biol. 78:11-21.

Elgin S.C.R. (1988). The formation and function of DNaseI hypersensitive sites in the process of gene activation. J. Biol. Chem., 263, pp. 19259–19262.

Espinoza C.R., Feeney A.J. (2005). The extent of histone acetylation correlates with the differential rearrangement frequency of individual VH genes in pro-B cells. J Immunol, 175:6668–75.

Fang W., Mueller D.L., Pennell C.A., Rivard J.J., Li Y.S., Hardy R.R., Schlissel M.S., Behrens T.W. (1996). Frequent aberrant immunoglobulin gene rearrangements in pro-B cells revealed by a bcl-xL transgene. Immunity 4:291-299.

Featherstone K., Wood A.L., Bowen A.J., Corcoran A.E. (2010). The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. J Biol Chem 285:9327–38.

Ferrier P., Krippl B., Blackwell T.K., Furley A.J., Suh H., Winoto A., Cook W.D., Hood L., Costantini F., Alt F.W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. EMBO J, 9:117–25.

Forrester W.C., Epner E., Driscoll M.C., Enver T., Brice M., Papayannopoulou T., Groudine M. (1990). A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev. 4:1637-1649.

Fraser P. (2006). Transcriptional control thrown for a loop. Curr Opin Genet Dev. 16:490–495.

Garrett, F. E., Emelyanov, A. V., Sepulveda, M. A., Flanagan, P., Volpi, S., Li, F., Loukinov, D.,Eckhardt, L. A., Lobanenkov, V. V., Birshtein, B. K. (2005). Chromatin architecture near a potential 30 end of the IgH locus involves modular regulation of histone modifications during B-cell development and in vivo occupancy at CTCF sites. Mol. Cell. Biol. 25(4), 1511–1525.

Gaszner, M., Felsenfeld, G. (2006). Insulators: Exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. 7, 703–713.

Giallourakis C. C., Franklin A., Guo C., Hwei-Ling C., Hye Suk Y., Gallagher M., Perlot T., Milena A., Murphy A., Macdonald L., Yancopoulos G., Alt F.W. (2010). Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. Proc. Natl Acad. Sci. USA 107, 22207–22212.

Giannini S.L., Singh M., Calvo C.F., Ding G., Birshtein B.K. (1993). DNA regions flanking the mouse Ig 3' alpha enhancer are differentially methylated and DNAase I hypersensitive during B cell differentiation. J Immunol. 150:1772–80.

Gillies, S. D., Morrison, S. L., Oi, V. T., Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 33(3), 717–728.

Gong S., Zheng C., Doughty M.L., Losos K., Didkovsky N., Schambra U.B., Nowak N.J., Joyner A., Leblanc G., Hatten M.E., Heintz N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature. 425:917–925.

Grawunder U., Leu T.M., Schatz D.G., Werner A., Rolink A.G., Melchers F., Winkler T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. Immunity. (3), 601-608.

Grosveld F., van Assendelft G.B., Greaves D.R., Kollias G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell. 51:975-985.

Gross D.S., Garrard W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem. 57:159–197.

Grundy G.J., Yang W., Gellert M. (2010) Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. Proc Natl Acad Sci USA, 107:22487–92.

Gu H., Zou Y.R., Rajewsky K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell. 73(6):1155-64.

Guo, C., Gerasimova T., Hao H., Ivanova I., Chakraborty T., Selimyan R., Oltz E.M., Sen R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell. 147, 332–343.

Guo C., Yoon H.S., Franklin A., Jain S., Ebert A., Cheng H.L., Hansen E., Despo O., Bossen C., Vettermann C., Bates J.G., Richards N., Myers D., Patel H., Gallagher M., Schlissel M.S., Murre C., Busslinger M., Giallourakis C.C., Alt F.W. (2011). CTCF-binding elements mediate control of V(D)J recombination. Nature. 477:424–30.

Gustavsson N., Lao Y., Maximov A., Chuang J.C., Kostromina E., Repa J.J., Li C., Radda G.K., Südhof T.C., Han W. (2008). Impaired insulin secretion and glucose intolerance in synaptotagmin-7 null mutant mice. Proc Natl Acad Sci U S A. 105(10):3992-7.

Heintzman N.D., Hon G.C., Hawkins R.D., Kheradpour P., Stark A., Harp L.F., Ye Z., Lee L.K., Stuart R.K., Ching C.W., Ching K.A., Antosiewicz-Bourget J.E., Liu H., Zhang X., Green R.D., Hesslein D.G., Schatz D.G. (2001). Factors and forces controlling V(D)J recombination. Adv Immunol. 78:169-232.

Hesslein, D. G., Pflugh, D. L., Chowdhury, D., Bothwell, A. L., Sen, R., Schatz, D. G. (2003). Pax5 is required for recombination of transcribed, acetylated, 50 IgH V gene segments. Genes Dev. 17(1), 37–42.

Ho Y., Elefant F., Cooke N., Liebhaber S. (2002). A defined locus control region determinant links chromatin domain acetylation with long-range gene activation. Mol Cell. 9:291-302.

Hu Q., Kwon Y.S., Nunez E., Cardamone M.D., Hutt K.R., Ohgi K.A., Garcia-Bassets I., Rose D.W., Glass C.K., Rosenfeld M.G., Fu X.D. (2008). Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. Proc Natl Acad Sci USA. Dec 9;105(49).

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. (2001). Nature, 409:860–921.

Jaeger S., Fernandez B., Ferrier P. (2013). Epigenetic aspects of lymphocyte antigen receptor gene rearrangement or 'when stochasticity completes randomness'. Immunology. 139(2):141-50.

Jeevan-Raj B.P., Robert I., Heyer V., Page A., Wang J.H., Cammas F., Alt F.W., Losson R., Reina-San-Martin B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. J Exp Med. 208(8):1649-60.

Jenuwein T., Forrester W.C., Fernandez-Herrero L.A., Laible G., Dull M., Grosschedl R. (1997) Extension of chromatin accessibility by nuclear matrix attachment regions. Nature. 385:269–272.

Jhunjhunwala S., van Zelm M.C., Peak M.M., Murre C. (2009). Chromatin architecture and the generation of antigen receptor diversity. Cell. 138(3):435-48.

Ji Y., Resch W., Corbett E., Yamane A., Casellas R., Schatz D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell. 141:419–31.

Ji Y., Little A.J., Banerjee J.K., Hao B., Oltz E.M., Krangel M.S., Schatz D.G. (2010). Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. J Exp Med, 207:2809–16.

John S., Sabo P.J., Canfield T.K., Lee K., Vong S., Weaver M., Wang H., Vierstra J., Reynolds A.P., Thurman R.E., Stamatoyannopoulos J.A. (2013) Genome-scale mapping of DNase I hypersensitivity. Curr Protoc Mol Biol. Jul;Chapter 27:Unit 21.27.

Johnson K., Chaumeil J., Skok J.A. (2010). Epigenetic regulation of V(D)J recombination. Essays Biochem 48:221–43.

Johnson K., Hashimshony T., Sawai C.M., Pongubala J.M., Skok J.A., Aifantis I., Singh H. (2008) Regulation of immunoglobulin light-chain recombination by the transcription factor IRF-4 and the attenuation of interleukin-7 signaling. Immunity. 28:335–45.

Johnston, C. M., Wood, A. L., Bolland, D. J., Corcoran, A. E. (2006). Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. J. Immunol. 176(7), 4221–4234.

Jones J.M., Gellert M. (2002). Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. EMBO J 21:4162–71.

Ju Z., Volpi S.A., Hassan R., Martinez N., Giannini S.L., Gold T., Birshtein B.K. (2007). Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3' regulatory region. J Biol Chem 282:35169–78.

Kabat E.A., Wu T.T., Perry H.M., Gottesman K.S., Foeller C. (1991). Sequences of Proteins of Immunological Interest. Bethesda, MD: U.S. Department of Health and Human Services.

Karasuyama, H., Kudo A., Melchers, F. (1990). The proteins encoded by the VpreB and lambda 5 pre-B cell-specific genes can associate with each other and with mu heavy chain. J. Exp. Med. 172, 969–972.

Kim U.J., Birren B.W., Slepak T., Mancino V., Boysen C., Kang H.L., Simon M.I., Shizuya H. (1996). Construction and characterization of a human bacterial artificial chromosome library. Genomics 34(2):213-8.

Kottmann, A. H., Zevnik, B., Welte, M., Nielsen, P. J., Kohler, G. (1994). A second promoter and enhancer element within the immunoglobulin heavy chain locus. Eur. J. Immunol. 24(4), 817–821.

Krebs J., Goldstein E., Kilpatrick S., (2012). Lewin's Genes XI, Jones & Bartlett Learning; 11th ed.

Kunitz, M. (1940). Crystalline Ribonuclease. J Gen Physiol. 24(1): 15–32.

Lahm A., Suck D. (1991). DNase I-induced DNA conformation: 2 A° structure of a DNase I-octamer complex. J. Mol. Biol. 222: 645–667.

Lazarovici A., Zhou T., Shafer A., Dantas Machado A.C., Riley T.R., Sandstrom R., Sabo P.J., Lu Y., Rohs R., Stamatoyannopoulos J.A., Bussemaker H.J. (2013) Probing DNA shape and methylation state on a genomic scale with DNase I. Proc Natl Acad Sci U S A. 110(16):6376-81.

Lazorchak A.S., Schlissel M.S., Zhuang Y. (2006). E2A and IRF-4/Pip promote chromatin modification and transcription of the immunoglobulin j locus in pre-B cells. Mol Cell Biol, 26:810–21.

Li Q., Harju S., Peterson K.R. (1999). Locus control regions: coming of age at a decade plus. Trends Genet.15:403–408.

Lieber M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem 79:181–211.

Lin C.Y., Vega V.B., Thomsen J.S., Zhang T., Kong S.L., Xie M., Chiu K.P., Lipovich L., Barnett D.H., Stossi F., Yeo A., George J., Kuznetsov V.A., Lee Y.K., Charn T.H.,

Palanisamy N., Miller L.D., Cheung E., Katzenellenbogen B.S., Ruan Y., Bourque G., Wei C.L, Liu E.T. (2007). Whole-genome cartography of estrogen receptor alpha binding sites. PLoS Genet. J 3(6).

Ling J.Q., Li T., Hu J.F., Vu T.H., Chen H.L., Qiu X.W., Cherry A.M., Hoffman A.R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312, 269–272.

Litt M.D., Simpson M., Recillas-Targa F., Prioleau M.N., Felsenfeld G. (2001). Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. EMBO J. 20(9):2224-35.

Liu Y., Subrahmanyam R., Chakraborty T., Sen R., Desiderio S. (2007). A plant homeodomain in RAG-2 that binds hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. Immunity. 27:561–71.

Lobanenkov V.V., Stewart R., Thomson J.A., Crawford G.E., Kellis M., Ren B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 459:108–112.

Loffert D, Ehlich A, Müller W, Rajewsky K. (1996). Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development. Immunity. (2):133-44.

Lomvardas S., Barnea G., Pisapia D.J., Mendelsohn M., Kirkland J., Axel R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell. 126(2):403-13.

Longo N.S., Lipsky P.E. (2006). Why do B cells mutate their immunoglobulin receptors? Trends Immunol 27(8): 374-380.

Lu J., Pazin M.J., Ravid K. (2004). Properties of ets-1 binding to chromatin and its effect on platelet factor 4 gene expression. Mol. Cell. Biol. 24:428–441.

Ma A., Fisher P., Dildrop R., Oltz E., Rathbun G., Achacoso P., Stall A., Alt F.W. (1992). Surface IgM mediated regulation of RAG gene expression in E mu-N-myc B cell lines. EMBO J. 11:2727-2734.

Maës J., Chappaz S., Cavelier P., O'Neill L., Turner B., Rougeon F., Goodhardt M. (2006). Activation of V(D)J recombination at the IgH chain JH locus occurs within a 6-kilobase chromatin domain and is associated with nucleosomal remodeling. J Immunol. 176(9):5409-17.

Malin S., Mcmanus S., Cobaleda C., Novatchkova M., Delogu A., Bouillet P., Strasser A., Busslinger, M. (2010). Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. Nat Immunol. 11:171–179.

Martensson I.L., Keenan R.A., Licence S. (2007). The pre-B-cell receptor. Curr Opin Immunol. 19(2): 137-142.

Matheson L.S., Corcoran A.E. (2012). Local and global epigenetic regulation of V(D)J recombination. Curr Top Microbiol Immunol. 356:65-89.

McArthur M., S. Gerum, G. Stamatoyannopoulos. (2001). Quantification of DNase 1 sensitivity by real-time PCR: quantitative analysis of DNase 1 hypersensitivity of the mouse beta-globin LCR. J. Mol. Biol. 313:27–34.

McMurry M.T., Krangel M.S. (2000). A role for histone acetylation in the developmental regulation of VDJ recombination. Science. 287:495–8.

Medina K.L., Singh H. (2005) Genetic networks that regulate B lymphopoiesis. Curr Opin Hematol. 12(3):203-9.

Medvedovic J., Ebert A., Tagoh H., Tamir I.M., Schwickert T.A., Novatchkova M., Sun Q., Huis In 't Veld P.J., Guo C., Yoon H.S., Denizot Y., Holwerda S.J., de Laat W., Cogné M., Shi Y., Alt F.W., Busslinger M. (2013). Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. Immunity 39:229–44.

Melchers F., ten Boekel E., Yamagami T., Andersson J., Rolink A. (1999). The roles of preB and B cell receptors in the stepwise allelic exclusion of mouse IgH and L chain gene loci. Semin Immunol. 11(5):307-17.

Michaelson J.S., Giannini S.L., Birshtein B.K. (1995) Identification of 3’ alpha-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. Nucleic Acids Res. 23:975–81.

Mills F.C., Fisher L.M., Kuroda R., Ford A.M., Gould H.J. (1983). DNase I hypersensitive sites in the chromatin of human mu immunoglobulin heavy-chain genes. Nature. 306:809–12.

Mito Y., Henikoff J.G., Henikoff S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet 37:1090–1097.

Mito, Y., Henikoff, J.G., Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science 315, 1408–1411.

Morshead K.B., Ciccone D.N., Taverna S.D., Allis C.D., Oettinger M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. Proc Natl Acad Sci USA, 100:11577–82.

Mueller-Storm H.P., Sogo J.M., Schaffner W. (1989). An enhancer stimulates transcription in trans when attached to the promoter via a protein bridge. Cell. 58(4):767-77.

Mundy C.L., Patenge N., Matthews A.G., Oettinger M.A. (2002). Assembly of the RAG1/RAG2 synaptic complex. Mol Cell Biol 22:69–77.

Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., Honjo T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102(5): 553-563.

Mutskov V., Felsenfeld G. (2009). The human insulin gene is part of a large open chromatin domain specific for human islets. Proc Natl Acad Sci USA 106: 17419–17424.

Navas P.A., Swank R.A., Yu M., Peterson K.R., Stamatoyannopoulos, G. (2003). Mutation of a transcriptional motif of a distant regulatory element reduces the expression of embryonic and fetal globin genes. Hum. Mol. Genet. 12(22): 2941–2948.

Neusser M., Schubel V., Koch A., Cremer T., Muller S. (2007). Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. Chromosoma. 116(3):307-20.

Neuberger, M.S. (1983). Expression and regulation of immunoglobulin heavy chain genes transfected into lymphoid cells. EMBO J 2: 1373-1378.

Nikolajczyk, B.S., W. Dang, R. Sen. (1999). Mechanisms of mu enhancer regulation in B lymphocytes. Cold Spring Harb. Symp. Quant. Biol. 64:99–107.

Nitschke, L., Kestler, J., Tallone, T., Pelkonen, S., Pelkonen, J. (2001). Deletion of the DQ52 element within the Ig heavy chain locus leads to a selective reduction in VDJ recombination and altered D gene usage. J. Immunol. 166(4), 2540–2552.

Nobrega, M.A., Ovcharenko, I., Afzal, V. Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. Science 302, 413.

Nutt, S.L., Heavey, B., Rolink, A.G., Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature. 401(6753): 556-562.

Nutt, S. L., Urba´nek, P., Rolink, A., Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: Difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. Genes Dev. 11(4), 476–491.

Oestreich K.J, Cobb R.M, Pierce S., Chen J., Ferrier P., Oltz E.M. (2006) Regulation of TCRbeta gene assembly by a promoter/enhancer holocomplex. Immunity. 24:381–91.

Oliveri M., Daga A., Lunardi C., Navone R., Millo R., Puccetti A. (2004). DNase I behaves as a transcription factor which modulates Fas expression in human cells. Eur J Immunol. 34(1):273–279.

Ong C.T., Corces V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet. 12(4):283-93.

Orphanides G., Reinberg D. (2000). RNA polymerase II elongation through chromatin. Nature 407:471–475.

Osborne C.S., Chakalova L., Mitchell J.A., Horton A., Wood A.L., Bolland D.J., Corcoran A.E., Fraser P. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol 5(192).

Osipovich O., Milley R., Meade A., Tachibana M., Shinkai Y., Krangel M.S., Oltz E.M. (2004). Targeted inhibition of V(D)J recombination by a histone methyltransferase. Nat Immunol, 5:309–16.

Palmiter R.D., Sandgren E.P., Koeller D.M., Brinster R.L. (1993). Distal regulatory elements from the mouse metallothionein locus stimulate gene expression in transgenic mice. Mol Cell Biol. 13:5266-5275.

Palstra R.J., de Laat W., Grosveld F. (2008). Beta-globin regulation and long-range interactions. Adv Genet. 61:107-42.

Pawlitzky, I., Angeles, C. V., Siegel, A. M., Stanton, M. L., Riblet, R., Brodeur, P. H. (2006). Identification of a candidate regulatory element within the 5 flanking region of the mouse Igh locus defined by pro-B cell-specific hypersensitivity associated with binding of PU.1, Pax5, and E2A. J. Immunol. 176(11), 6839–6851.

Peitsch M.C., Irmler M., French L.E., Tschopp J. (1995). Genomic organisation and expression of mouse deoxyribonuclease I. Biochem Biophys Res Commun. 207(1):62-8.

Perlot, T., Alt, F. W., Bassing, C. H., Suh, H.,Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc. Natl. Acad. Sci. USA 102(40), 14362–14367.

Pettersson S., Cook G.P., Bruggemann M., Williams G.T., Neuberger M.S. (1990). A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus. Nature. 344:165–8.

Phillips, J. E., Corces, V. G. (2009). CTCF: master weaver of the genome. Cell. 137:1194–1211.

Purizaca J., Meza I., Pelayo R. (2012). Early lymphoid development and microenvironmental cues in B-cell acute lymphoblastic leukemia. Arch Med Res. 43(2):89-101.

Rajewsky, K. (1996). Clonal Selection and learning in the antibody system. Nature. 381:751-758.

Ratajczak M.Z., Kim C., Ratajczak J., Janowska-Wieczorek A. (2013). Innate immunity as orchestrator of bone marrow homing for hematopoietic stem/progenitor cells. Adv Exp Med Biol. 735:219-32.

Reth M.G., Alt F.W. (1984). Novel immunoglobulin heavy chains are produced from DJH gene segment rearrangements in lymphoid cells. Nature. 312:418–423.

Retter, I., Chevillard, C., Scharfe, M., Conrad, A., Hafner, M., Im T.H., Ludewig, M., Nordsiek, G., Severitt, S., Thies, S., Mauhar, A., Blocker, H., Muller, W., Riblet, R. (2007). Sequence and characterization of the Ig heavy chain constant and partial variable region of the mouse strain 129S1. J Immunol 179(4): 2419-2427.

Revilla-I-Domingo R., Bilic I., Vilagos B., Tagoh H., Ebert A., Tamir I.M., Smeenk L., Trupke J., Sommer A., Jaritz M., Busslinger M. (2012). The B-cell identity factor Pax5

regulates distinct transcriptional programmes in early and late B lymphopoiesis. EMBO J. 31(14):3130-46.

Roth D.B., Zhu C., Gellert M. (1993). Characterization of broken DNA molecules associated with V(D)J recombination. Proc Natl Acad Sci U.S.A. 90:10788–92.

Rouaud P., Vincent-Fabert C., Fiancette R., Cogne M., Pinaud E., Denizot Y. (2012). Enhancers located in heavy chain regulatory region (hs3a, hs1,2, hs3b, and hs4) are dispensable for diversity of VDJ recombination. J Biol Chem. 287:8356–60.

Sadofsky M.J., Hesse J.E., McBlane J.F., Gellert M. (1993). Expression and V(D)J recombination activity of mutated RAG-1 proteins. Nucleic Acids Res. 21:5644–50.

Sagai T., Amano T., Tamura M., Mizushina Y., Sumiyama K., Shiroishi T. (2009) A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. Development 136:1665–1674

Sakai, E., Bottaro, A., Davidson, L., Sleckman, B. P., Alt, F. W. (1999). Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. Proc. Natl. Acad. Sci. U.S.A. 96(4), 1526–1531.

Schatz D.G., Swanson P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu Rev Genet. 45:167-202.

Schlissel M., Constantinescu A., Morrow T., Baxter M., Peng A. (1993). Double-strand signal sequence breaks in V(D)J recombination are blunt, 5′-phosphorylated, RAG-dependent, and cell cycle regulated. Genes Dev. 7:2520–32.

Schlissel M.S. (1994). Morrow T. Ig heavy chain protein controls B cell development by regulating germ-line transcription and retargeting V(D)J recombination. J Immunol. 153:1645-1657.

Schlissel, M., Voronova, A., Baltimore, D. (1991). Helix-loop-helix transcription factor E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line. Genes Dev. 5(8), 1367–1376.

Schubeler D., Groudine M., Bender M.A. (2001). The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. Proc Natl Acad Sci U.S.A. 98:11432–7.

Sekimata M., Perez-Melgosa M., Miller S.A., Weinmann A.S., Sabo P.J., Sandstrom R., Dorschner M.O., Stamatoyannopoulos J.A., Wilson C.B. (2009). CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-g locus. Immunity 31: 551–564.

Sen R., Oltz E. (2006). Genetic and epigenetic regulation of IgH gene assembly. Curr Opin Immunol.18(3):237-42.

Sexton T., Bantignies F., Cavalli G. (2009). Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. Semin Cell Dev Biol. 20:849–855.

Sharan S.K., Thomason L.C., Kuznetsov S.G., Court D.L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. 4(2):206-23.

Shimazaki N., Tsai A.G., Lieber M.R. (2009). H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol Cell. 34:535–44.

Sheffield, N. C., Thurman, R. E., Song, L., Safi, A., Stamatoyannopuolos, J. A., Lenhard, B.. Crawford G.E., Furey T.S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. 23. 777–788.

Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y., Simon M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U.S.A. 89(18):8794-7.

Silver D.P., Spanopoulou E., Mulligan R.C., Baltimore D. (1993). Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. Proc Natl Acad Sci U.S.A. 90:6100–4.

Simpson E.M., Linder C.C., Sargent E.E., Davisson M.T., Mobraaten L.E., Sharp J.J. (1997). Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. Nat Genet. 16(1):19-27.

Sperling K., Kerem B.S., Goitein R., Kottusch V., Cedar H., Marcus M. (1985). DNaseI sensitivity in faculative and constituitive heterochromatin. Chromosoma, 93, pp. 38–42.

Spicuglia S., Kumar S., Yeh J-H., Vachez E., Chasson L., Gorbatch S., Cautres J., Ferrier P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol Cell. 10:1479–87.

Spilianakis C.G., Lalioti M.D., Town T., Lee, G.R., Flavell R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature. 435, 637–645.

Spilianakis C.G., Flavell R.A. (2006). Molecular biology. Managing associations between different chromosomes. Science. 312(5771):207-8.

Splinter E., Heath H., Kooren J., Palstra R.J., Klous P., Grosveld F., Galjart N., de Laat W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 20(17):2349-54.

Spolski R., Miescher G., Erard F., Risser R., MacDonald H.R., Mak T.W. (1988). Regulation of expression of T cell gamma chain, L3T4 and Ly-2 messages in Abelson/Moloney virus-transformed T cell lines. Eur J Immunol. 18(2):295-300.

Stanhope-Baker P., Hudson K.M., Shaffer A.L., Constantinescu A., Schlissel M.S. (1996). Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. Cell. 85:887–97.

Strauss W.M. (2001). Preparation of genomic DNA from mammalian tissue. Curr Protoc Immunol. 2001 May;Chapter 10:Unit 10.2.

Su L.K., Kadesch T. (1990). The immunoglobulin heavy-chain enhancer functions as the promoter for I mu sterile transcription. Mol. Cell. Biol. 10(6), 2619–2624.

Subrahmanyam R., Du H., Ivanova I., Chakraborty T., Yanhong J., Zhang Y., Alt F.W., Schatz D.G., Sen R., (2012). Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. Nat Immunol. 13:1205–12.

Subrahmanyam R., Sen R. (2012). Epigenetic features that regulate IgH locus recombination and expression. Curr Top Microbiol Immunol. 356:39-63.

Suck D. (1994). DNA recognition by DNase I. J Mol Recognit. 7(2):65-70.

Tolhuis B., Palstra R.J., Splinter E., Grosveld F., de Laat W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell.10(6):1453-65.

Tsukada S., Sugiyama H., Oka Y., Kishimoto S. (1990). Estimation of D segment usage in initial D to JH joinings in a murine immature B cell line. Preferential usage of DFL16. 1, the most 5' D segment and DQ52, the most JH-proximal D segment. J Immunol. 144:4053–9.

Valjent E., Bertran-Gonzalez J., Hervé D., Fisone G., Girault J.A. (2009). Looking BAC at striatal signaling: cell-specific analysis in new transgenic mice. Trends Neurosci. 32(10):538-47.

van Gent D.C., Ramsden D.A., Gellert M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell. 85:107–13.

Vaquerizas J.M., Kummerfeld S.K., Teichmann S.A., Luscombe N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 10:252–263.

Vieira P., Cumano A. (2004). Differentiation of B lymphocytes from hematopoietic stem cells. Methods Mol Biol. 271:67-76.

Volpi S.A.,Verma-Gaur J., Hassan R., Ju Z., Roa S., Chatterjee S., Werling U., Hou H. Jr., Will B., Steidl U., Scharff M., Edelman W., Feeney A.J., Birshtein B.K. (2012). Germline deletion of Igh 3' regulatory region elements hs 5, 6, 7 (hs5-7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. J Immunol. 188:2556–66.

Wang G., Dhar K., Swanson P.C., Levitus M., Chang Y. (2012). Real-time monitoring of RAG catalyzed DNA cleavage unveils dynamic changes in coding end association with the coding end complex. Nucleic Acids Res. 40:6082–96.

Wang L., Wuerffel R., Feldman S., Khamlichi A.A., Kenter A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. J Exp Med. 206(8):1817-30.

Welinder E., Ahsberg J., Sigvardsson M. (2011). B-lymphocyte commitment: identifying the point of no return. Semin Immunol. 23(5):335-40.

Weston S.A., Lahm A., Suck, D. (1992). X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 A° resolution. J. Mol. Biol. 226: 1237–1256.

Whitehurst C.E., Chattopadhyay S., Chen J. (1999). Control of V(D)J recombinational accessibility of the Db1 gene segment at the TCR b locus by a germline promoter. Immunity. 10:313–22.

Whitehurst C.E., Schlissel M.S., Chen J. (2000). Deletion of germline promoter PD b1 from the TCR b locus causes hypermethylation that impairs D b1 recombination by multiple mechanisms. Immunity. 13:703–14.

Wuerffel R., Wang L., Grigera F., Manis J., Selsing E., Perlot T., Alt F.W., Cogne M., Pinaud E., Kenter A.L. (2007). S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. Immunity 27: 711–722.

Xu Z., Wei G., Chepelev I., Zhao K., Felsenfeld G. (2011). Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. Nat Struct Mol Biol. 18(3):372-8.

Yancopoulos G.D., Alt F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. Cell. 40(2):271–281.

Yang X.W., Model P., Heintz N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol. 15:859–865.

Ye, J. (2004). The immunoglobulin IGHD gene locus in C57BL/6 mice. Immunogenetics. 56(6):399-404.

Zhao Z., Tavoosidana G., Sjolinder M., Gondor A., Mariano P., Wang S., Kanduri C., Lezcano M., Sandhu K.S., Singh U., Pant V., Tiwari V., Kurukuti S., Ohlsson R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and Interchromos omal interactions. Nat Genet. 38:1341–1347.

Zhang Z., Espinoza C.R., Yu Z., Stephan R., He T., Williams G.S., Burrows P.D., Hagman J., Feeney A.J., Cooper M.D. (2006). Transcription factor Pax5 (BSAP) transactivates the RAG-mediated V(H)-to-DJ(H) rearrangement of immunoglobulin genes. Nat Immunol. 7:616–624.

Zhou, J., Ashouian, N., Delepine, M., Matsuda, F., Chevillard, C., Riblet, R., Schildkraut, C.L., Birshtein, B.K. (2002). The origin of a developmentally regulated Igh replicon is located near the border of regulatory domains for Igh replication and expression. Proc Natl Acad Sci U.S.A. 99(21): 13693-13698.

CURRICULUM VITAE

Name: Jonathon D. Bennett **Date:** September 29, 2014

Educational History:

Ph.D. expected	2014	Immunology Graduate Program Johns Hopkins School of Medicine Mentor: Dr. Ranjan Sen
B.A.	2003	Ancient Languages Wheaton College

Other Professional Experience:

Research Rotation	2008	Lab of Dr. Ranjan Sen National Institutes of Health National Institute of Aging
Research Rotation	2008	Lab of Dr. Lieping Chen Johns Hopkins Medical School
Research Rotation	2008	Lab of Dr. Elizabeth Jaffee Johns Hopkins Medical School
Research Rotation	2007	Lab of Dr. Jonathan Pevsner Johns Hopkins Medical School
Research Assistant	2003-2007	Johns Hopkins Medical School Department of Pathology

Publications:

Schütz C., Zoso A., Peng S., **Bennett J**., Schneck, J., Oelke, M. MHC-Ig induces memory T cell formation in vivo and inhibits tumor growth. Immunity, Inflammation and Disease (In Press).

Abstracts and Posters:

Bennett J., Ghosh A., Du H., Sen R. (2013). Enhancer Induced Changes In Chromatin Structure & Function. (NIH FARE).

Bennett J., Banerjee H., Ghosh A., Du H., Sen R. (2012). Chromatin Structural Changes Induced by the Tissue-Specific Eμ Enhancer. (NIH Research Festival).

Bennett J., Guo C., Gerasimova T. Ivanova I., Selimyan R., Ghosh A., Du H., Sen R. (2011). Effect of enhancer location on chromatin structure, gene

transcription, and VDJ recombination for the murine IgH Locus. (NIH Young Investigator Conference).

Service and Leadership:

2013-2014	Johns Hopkins Medical Entrepreneurship Club
2009-2013	President – C.A.R.E. Community Association